

Genomics and taxonomy for all

Principles of access to public and private databases are often contentious, and a proposal in this issue will no doubt spark more debate. Meanwhile, *Nature* is taking a small step towards a database for taxonomists.

Arguments about access to published data reached a peak of intensity when the draft sequence of the human genome was published in February last year. They resurface from time to time, and this week Ari Patrinos and Dan Drell make the case for allowing industry to restrict access as a trade-off for making valuable data publicly available (see page 589). Increased public access to private data is desirable, but *Nature* continues to believe that restricted access to data that we publish is in general inappropriate (see *Nature* 409, 745; 2001), particularly where public projects and databases exist. But it is surely right that the research community should consider these alternative proposals, and we encourage responses.

Many areas of biology have community databases and others are being developed for microarrays and brain imaging, for example. However, there is one core area of biology that is too often overshadowed but that also needs to take steps to provide greater access to its immense store of knowledge and annotation: taxonomy, the formal nomenclature and description of organisms.

Without taxonomy, nobody would be sure of the identity of the organisms they were interested in, or whether they belonged to the same or different species as the organisms studied by others. Without taxonomy, there would be no meaningful genome projects, and medical science, for one, would be seriously compromised. Without taxonomy, there could be no systematics, the related but distinct business of arranging species' names into an order that reflects their evolutionary relationships. Without taxonomy, we could not begin to understand biodiversity and the related issue of conservation. For a variety of reasons, some of them self-inflicted — recently explored

by Charles Godfray (*Nature* 417, 17–19; 2002) — taxonomists have a poor image among other biologists. Taxonomy is starved of funds, whereas the arms of biology that rely fundamentally on it attract both money and publicity.

The great fragmentation of taxonomic publication has contributed to taxonomy's parlous state. Taxonomy would benefit from a high-profile, centralized repository of nomenclature. *Nature* is now taking a small step towards that end, requiring that authors of papers featuring new taxonomy should file this information with a recognized institution. We have set up such an arrangement with the Linnean Society of London, which is the oldest body in the world concerned with taxonomy, and which maintains the library and collections of Carl Von Linné (Linnaeus), who founded the modern system of taxonomy in the eighteenth century. From 1 August 2002, the authors of any paper containing the formal nomenclature and description of species that has been accepted in principle by *Nature*, shall be required as a condition of acceptance to send a preprint to the Linnean Society of London, Burlington House, Piccadilly, London W1J 0BF, UK or an electronic version by e-mail to john@linnean.org. Deposition shall be voluntary for papers accepted in principle before 1 August 2002.

Another ill that besets taxonomy is the inability of taxonomists to forge united initiatives. This is why our action as a journal is unilateral. Nevertheless, it is our hope that other journals will adopt the same policy, encouraging the future development of an instantly accessible electronic archive with agreed standards. And, as with genome sequence databases, if several recognized institutions decide to host taxonomy databases that *Nature* can support, so much the better. ■

Remove barricades! Preserve culture!

Congratulations to French publications for their pragmatic approach to language. Other institutions should follow suit.

President John F. Kennedy struck the right note when in 1963 he famously endorsed freedom by declaring “*Ich bin ein Berliner*”.

But foraying into a foreign language carries its perils, confronted daily by many of the millions of scientists for whom English is not the mother tongue. The English *lingua franca* of science is often resented as Anglo-Saxon cultural imperialism; a recent cartoon in a French newspaper summed up this sentiment portraying George W. Bush on his European voyage as “*Ich bin ein Hamburger*”.

Every language is rich, with unique and untranslatable treasures of vernacular and elegance. No surprise that the British royal family continues to express its motto “*Honni soit qui mal y pense*” in the language of Molière rather than the clumsy translation “Shame to him who thinks ill of it”. Or take *Petits Débrouillards*, a scheme to teach French children hands-on science. It could translate as ‘little inventors’ or even ‘little smart-arses’, but English cannot capture its sense of *bricolage* encapsulating simultaneously notions of backyard science, improvisation and invention.

Leaving aside its sometimes farcical attempts to prevent the anglicisms that are part of natural gene flow between languages, the Francophone world — 160 million people in 49 different countries —

in particular deserves *félicitations* for its historic determination to preserve its language and culture. Is it now capitulating? Last month, *Le Monde* created *une tempête* by publishing a weekly supplement of articles from *The New York Times* and, *s'il vous plaît*, in English. This week, the Académie des Sciences decided that its *Comptes Rendus* would in future give “a preference” to articles in English (see page 581). But these moves are simply *pragmatique*. *Le Monde* rightly argues that its supplement gives readers a different perspective on world events, that the writings of *The New York Times* journalists are best expressed in their original language, and that over half its readers understand English. The academy is simply acknowledging *la réalité* that using English is the only way to get read more widely.

Such linguistic *pragmatisme* should not displace expressive and subtle discussion about science in native languages. But it needs to permeate all research organizations wishing to attract international talent. It is encouraging that many of Europe's young scientists now speak two or more languages. The British are the continent's laggards: over two-thirds speak *pas un mot d'une langue étrangère*. Irritatingly to non-Anglophones, UK and US scientists don't suffer for such *philistinisme*: the international spoken language of science is English, however broken. ■





Savings bank
World Bank sets out
plan to conserve
archaeological sites
p577



Galapagos slick
Ruined iguana study
leads ecologist to sue
owners of oil tanker
p578



Making a splash
Space and ocean
researchers pool
their resources
p579



Washed out
NASA water-cycle
satellite springs
a leak
p580

Alpine detector fails to confirm Italian sighting of dark matter

Alison Abbott, Munich

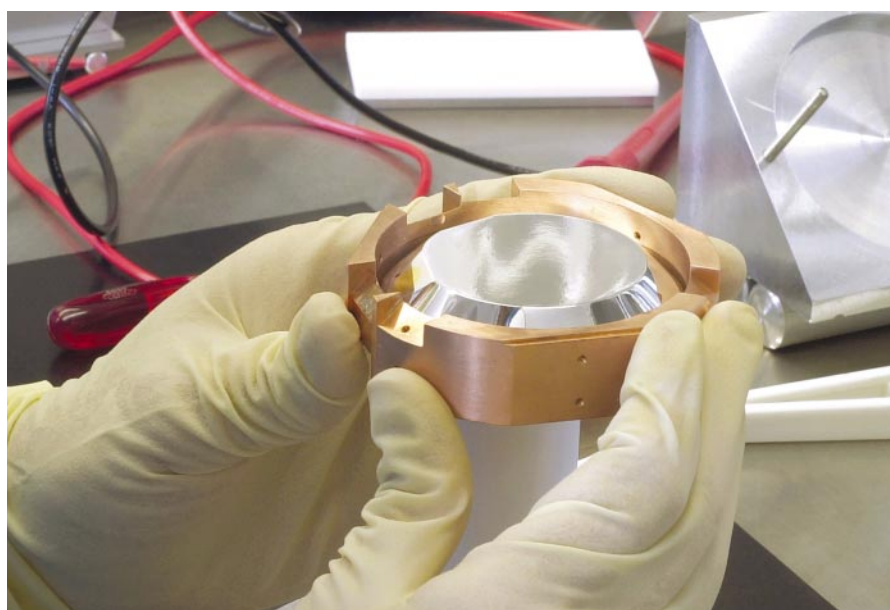
Despite a three-month search for the mysterious particles thought to make up 90% of the Universe's mass, dark matter has remained rooted in its shadowy world.

French high-energy physicists last week revealed that their highly sensitive detector had failed to detect a single particle of dark matter with the characteristics claimed to have been spotted by an Italian team two years ago.

The results, presented at a hastily arranged special session of the 20th International Conference on Neutrino Physics and Astrophysics in Munich, contradict findings of the four-year DAMA experiment, led by Rita Bernabei of the University of Rome.

Bernabei's team controversially claimed that it had detected WIMPs — weakly interacting massive particles — a theoretical type of dark-matter particle (see R. Bernabei *et al.* *Eur. Phys. J. C* **18**, 283–292; 2000).

DAMA, based in the Gran Sasso laboratory near Rome, uses a detector containing iodine, which measures the light emitted when incoming WIMPs collide with the iodine nuclei. The change in the Earth's orientation as it orbits the Sun means that its interaction with the cloud of WIMPs that is believed to fill the Solar System should vary on a seasonal basis. Bern-



The highly sensitive detectors at Edelweiss use germanium to spot any interactions with dark matter.

abei claimed to have detected this variation.

DAMA is isolated from interfering radiation by a covering of 1,400 metres of rock. But many physicists claimed that the detector was not sensitive enough to distinguish a true WIMP response from background noise. In addition, most sources of noise,

such as cosmic rays, would be subject to seasonal variation, they argued.

Two years ago, the CDMS experiment, led by Bernard Sadoulet, a cosmologist at the University of California, Berkeley, failed to find conclusive evidence for the type of WIMP that Bernabei claimed to have

Supreme Court closes loophole for copycat patents

Jonathan Knight, San Francisco

Holders of valuable patents breathed a sigh of relief last week, when the US Supreme Court reversed a lower-court decision that many feared would pave the way for more copycat inventions.

In 2000, a federal appeals court upset supporters of tight patent protection by limiting the 'doctrine of equivalents'. This principle broadly states that a patent on blue nails can be extended to cover red nails as well, as their function is essentially the same.

In a decision known as the Festo ruling, the federal court had decreed that if a patent's claim had been narrowed during review —

say from yellow and red nails to red nails only — it could not be widened again later by claiming equivalence. The case involved Festo Corporation, a machine-tool-maker in New York state, which argued that its patent on a part for a robotic arm should cover a similar device made by a Japanese rival.

As most patents are the product of negotiation with the Patent and Trademark Office, patent holders were worried that Festo would open the floodgates to close imitations of patented inventions.

In one related case, for example, MedImmune of Gaithersburg, Maryland, is fighting with the British biotechnology

company Celltech to claim rights to an antibody for treating respiratory infections. MedImmune's version differs by one amino acid in 1,320 from Celltech's patented one.

One of the several petitions MedImmune filed with the Supreme Court argues that the doctrine of equivalents has been too widely interpreted, and should be limited as the federal court suggested.

But more than a dozen universities and research organizations, including Stanford University and the Massachusetts Institute of Technology, filed briefs urging the Supreme Court to undo the court judgement, which it did on 28 May.

detected. But the CDMS detector is only around 10 metres below Stanford University's campus, and background noise was too high for researchers to draw unambiguous conclusions.

The latest results were obtained by researchers working on the Edelweiss detector at the Modane Underground Laboratory, located 1,700 metres below the Alps.

Edelweiss has lower levels of background noise than the CDMS detector. And, by virtue of using germanium instead of iodine, it is more sensitive than DAMA. Germanium allows the measurement of two properties of any potential interactions — temperature change and an ionization signal — rather than just one.

"The results reported by the DAMA team could have been generated by background events or could have been statistical fluctuations," says Gilles Gerbier of the Saclay Centre of the French Atomic Energy Commission near Paris, and a member of the Edelweiss collaboration.

Other researchers agree. "The DAMA results were hard to believe, but also hard to disprove. The Edelweiss results are very convincing," says Yorck Ramachers, who works on another dark-matter experiment at Gran Sasso. But Bernabei is sticking to her guns, insisting that the DAMA WIMPs are real, and that experiments using different detectors are not comparable.

The search for WIMPs with other characteristics will go on. Edelweiss will be able to extend its search after two more detectors are installed in the next nine months — a total of 21 will be online by the end of 2003. In addition, the CDMS experiment will soon move to a deeper site in north Minnesota. ■

♦ <http://edelweiss.in2p3.fr>

South African cabinet backs merger plan for universities

Michael Cherry, Cape Town

South Africa is to proceed with a controversial plan to merge institutes of higher education that taught blacks and whites separately under apartheid (see *Nature* 417, 377–378; 2002).

The cabinet accepted the merger plan late last month after Kader Asmal, the education minister, made concessions that will allow two of the nation's best-known historically black universities — the University of the Western Cape and the University of Fort Hare — to retain their separate identities.

The country's 21 universities and 15 technikons (polytechnics) will now be streamlined into 11 universities, 6 technikons and 4 comprehensive institutions that will offer both university and technikon programmes. The government says that this will create "a system that is equitable, academically and financially sustainable, and productive".

The government also hopes to counter falling registration numbers by increasing the participation rate — the percentage of 20–24-year-olds enrolled in higher education — from 15% to 20% over the next 10 years. This will involve an estimated additional 200,000 students, most of whom will need to be supported by increased government allocations to the National Student Financial Aid Scheme — although the government also hopes to recruit more students from other countries in the region.

The South African government is also committed to reducing the percentage of



The historically black University of the Western Cape is set to retain its separate identity.

UWC

enrolments in the humanities from 49% to 40% over the next five to ten years, while increasing those in business or commerce from 26% to 30%, and those in science, engineering and technology from 25% to 30%.

Derrick Swartz, rector of the University of Fort Hare, said in a statement that the cabinet's decision had vindicated its long campaign against the merger, and that he hoped it would make the "new Fort Hare the major player in the region". But others were less satisfied. "These are senseless proposals that amount to merely tinkering with the status quo," Itumeleng Mosala, Rector of the Technikon North-West and chairman of the Association of Historically Disadvantaged Institutions, told the Johannesburg daily *Business Day*. ■

Chinese researcher accused of stealing cell samples

Rex Dalton, San Diego

A Chinese-born scientist has been held in a Californian jail for nearly three weeks, after being arrested for allegedly stealing materials and methods used to grow corneal cells which, police files claim, he intended to export to China.

Veterinary researcher Bin Han was jailed on 17 May after police searched his home and found records of stem-cell experiments and serum samples allegedly misappropriated from the University of California, Davis, where he is employed. They also found a ticket for a round-trip to China for the following week.

The jailing of Han — who was born in China but has been a US citizen since 1999 — is the latest example of the increasingly hard line US authorities are taking with researchers who stand accused of exporting

biological samples or information without permission. Last May, for example, two Japanese researchers were charged with industrial espionage for their alleged involvement in such exports (see *Nature* 411, 225–226; 2001).

At a state court hearing on 4 June in a suburb of Sacramento, California, a judge allowed Han to be released after authorities had taken possession of his passports. Han now faces a preliminary hearing on 16 July on at least one felony charge of embezzlement.

Since 1990, Han has worked in various laboratories at the University of California, Davis, after coming to the United States in 1989. During one period, according to university records, he also said he was running an investment firm on behalf of Chinese financiers.

Two years ago, Han became a postdoctoral fellow in the laboratory of ophthalmologist Ivan Schwab and dermatologist Rivkah Isseroff — who are investigating how to make stem cells grow into corneal epithelial cells to replace damaged eye tissue. The technique might also have wide application for making other epithelial cells, researchers say.

Early last month, court records show, officials at the University of California, Davis, became suspicious of Han after learning that he was setting up a stem-cell research laboratory in China, where he frequently travelled. Police raided Han's home after vials of human sera used in Schwab and Isseroff's research went missing.

If convicted, Han could face a prison term. Neither he nor his attorney could be reached for comment. ■

World Bank cracks down on cultural damage

Rex Dalton, Tucson

The World Bank is close to adopting a rigorous new policy that aims to safeguard archaeological and anthropological sites that could be affected by its development projects.

A policy drafted by bank officials calls for early assessment of cultural heritage resources in areas to be developed, and requires the borrowing nation to address the need to preserve physical resources. It also provides for emergency measures to conserve valuable sites.

Within a month, the board of the bank — one of the world's main backers of development projects — is expected to vote on the draft, which was agreed after three years of consultation with scientists and other interested parties.

Bank officials also hope that its new stance will prompt other agencies that fund development projects to embrace stronger policies to protect cultural heritage.

"The policy is potentially revolutionary," says Steven Brandt, a University of Florida archaeological anthropologist who sits on the bank's eight-person scientific advisory panel. "The bank is really sincere about this," he says.

Researchers hope the policy will also lead to training and placement of archaeologists and anthropologists in the countries where development projects are going forward. Many poor nations, especially in Africa, have few such scientists of their own.

But while the World Bank moves to adopt its new policy, researchers at last month's biennial meeting of the Society of Africanist Archaeologists in Tucson, Arizona, reported that significant problems persist with development projects now in progress.

In Ethiopia, for instance, Brandt says that scientists and volunteers are rushing to complete a last-minute cultural heritage assessment of an area near Jimma — where a reservoir is soon to be filled as part of the \$300-million Gilgel Gibe hydroelectric dam project, funded by the World Bank and the European Union (EU).

Brandt says that EU-funded roads for the project have been bulldozed through ancient cemeteries, infuriating local people and scientists. EU and bank officials declined to comment.

Scientists have only been able to muster about \$200,000 for surveys of the reservoir and surrounding area in southwest Ethiopia,



Time trial: archaeologists at Gilgel Gibe, Ethiopia, are racing to survey the site before a dam is built.

according to Brandt. "It has been a real scramble and a nightmare," he says.

♦ worldbank.org

♦ www.nilebasin.org

GILGEL GIBE ARCHAEOLOGICAL PROJECT

British science champion quits post

David Adam, London

A body charged with improving science communication in Britain is looking for new leadership after its illustrious chairperson resigned, complaining angrily about its lack of progress in meeting its goal.

Bridget Ogilvie, former head of the Wellcome Trust, the world's largest medical research charity, accused the Royal Society of blocking her attempts to turn the body, called Copus, into an umbrella organization for the various UK groups involved in science communication.

"To become an umbrella organization it has to be an organization where not one body dominates, but the Royal Society runs Copus and is not prepared to allow Copus to evolve," Ogilvie said. "Frankly I've got better things to do with my time." She announced her decision to surprised colleagues at the organization's spring council meeting on 17 May.

Formerly known as the Committee on the Public Understanding of Science, Copus was launched in 1985 by the Royal Society, the Royal Institution and the British Association. It was remodelled last year to include representatives from other organizations, including scientific societies, museums and the media, but effectively remains a branch of



Bridget Ogilvie: move stunned colleagues.

the Royal Society, which applies for government funds on its behalf and provides offices.

Ogilvie says she wanted Copus to evolve into a more democratic body in which member organizations are each able to exert roughly equal influence. But Stephen Cox, executive secretary of the Royal Society and a member of the Copus council, says that would not work because groups interested in science communication don't have the cash to support Copus in that way. "Currently we provide Copus with its home and infrastructure and as soon as you spin it off, all those costs have to be found from somewhere else," he says.

Other Copus council members question the need for an 'umbrella' organization. "We're all too busy with our own agendas to stand around keeping the rain off each other," says Jon Burch, executive secretary of the Royal Academy of Engineering.

Uncertainty ends for US nuclear lab as new leader is named

Rex Dalton, San Diego

A new director has been appointed at the Lawrence Livermore National Laboratory (LLNL) in California — ending weeks of uncertainty over the future leadership of the nuclear-weapons research facility.

Physicist Michael Anastasio, currently LLNL's deputy director for strategic operations, was approved as director on 4 June by the regents of the University of California (UC), which manages the lab for the US Department of Energy.

Anastasio has worked at LLNL for 22 years, and is said to have played a key role in maintaining the safety and reliability of nuclear weapons in the absence of testing. He will take over from the current director, Bruce Tarter, on 1 July. LLNL has 7,500 employees and an annual budget of around \$1.5 billion.

In April, Anastasio was one of a shortlist of four candidates when the UC regents almost named nuclear engineer Ray Juzaitis, an associate director at Los Alamos National Laboratory in New Mexico, as LLNL director. But that plan was abandoned, reportedly after some LLNL researchers voiced opposition to it (see *Nature* 417, 3; 2002).

Ecologist sues over wrecked iguana study

John Whitfield, London

An ecologist whose studies of marine iguanas in the Galapagos Islands were damaged by an oil spill is suing the oil tanker's owner and insurer.

Martin Wikelski of Princeton University in New Jersey is seeking US\$600,000 in damages as part of a wider, \$14-million claim being made by the Galapagos National Park against the master, owners and insurers of the *Jessica*, a tanker that ran aground on the island of San Cristóbal in January 2001. The park is also suing Petrocomercial, an arm of the Ecuadorian state oil company, which owned the oil.

Wikelski has been studying the Galapagos' marine iguanas since 1987. His work has revealed, for example, how the animals' bodies shrink in response to El Niño events (see *Nature* 403, 37–38; 2000). But after the oil spill, the focus of his observations switched sharply from evolution to pollution.

Following the accident, Wikelski and his colleagues monitored the iguanas at an oil-affected site and at one that had escaped



Black death: oil from the *Jessica* has killed a large number of marine iguanas on San Cristóbal island.

contamination. In this week's *Nature*, they report that by December 2001 more than 60% of the marked iguanas at the contaminated site had died, compared with none at the site that escaped pollution (*Nature* 417, 607–608; 2002).

The team suggests that small amounts of

oil killed microbes in the iguanas' guts that enable the animals to digest algae, causing them to starve. "Even a tiny amount of oil in the water can be very harmful," says Wikelski.

The spill terminated Wikelski's earlier experiments, and the \$600,000 represents some of the funding that Wikelski and his team had received for them from Germany's Max Planck Society, the University of Princeton and the US National Science Foundation.

The rest of the \$14-million claim covers the cost of cleaning up the spill and monitoring its effects, loss of tourism and fishing revenue, and damage to the Galapagos' pristine image. "We're still feeling the economic effects," says park biologist Mauricio Velasquez.

The case was filed with the High Court at Guayaquil, Ecuador, where hearings began this February. Terra Nova, the London-based insurer of the *Jessica*, is contesting the right of the Ecuadorian courts to hear it. The insurer also says that the *Jessica* was uninsured when it ran aground because the ship's owners had not paid their premiums and had not complied with the terms of their policy. "We're just not involved at all," says Andrew Barker, an underwriter with the company. Petrocomercial declined to comment on the case.

But Ider Valverde, the leader of the park's legal team, says that the insurer is liable because it failed to notify the authorities that the *Jessica* was no longer covered, as is required by Ecuadorian law.

"We have made it very clear that the park is a non-profit organization, and have tried several times to settle with the insurers in an amicable fashion," says Valverde. "I imagined that they would have preferred a nice quiet settlement, but they have decided to litigate instead." If he wins, Wikelski says that any damages will be used to set up a fund to support young Ecuadorian scientists.

Merger plans rattle bones in Berlin

Johanna Schwarz, Munich

Palaeontologists at Berlin's Museum of Natural History are up in arms over a reform plan that they fear could leave the dinosaur discipline on the edge of local extinction.

An outside panel of experts recommended in April that the institutes of palaeontology, mineralogy and zoology at the museum should merge into a single department, with one director.

But palaeontologists at the institute claim that the panel's report belittled their discipline. The report describes palaeontology as having "no scientific autonomy". It adds that the discipline lacks a theoretical approach as it is derived from the objects it studies and relies on "expeditions and preparation techniques".

"Palaeontology is definitely hard science," says Rainer Schoch, one of 15 palaeontologists at the museum, who says he finds the report's findings "unacceptable".

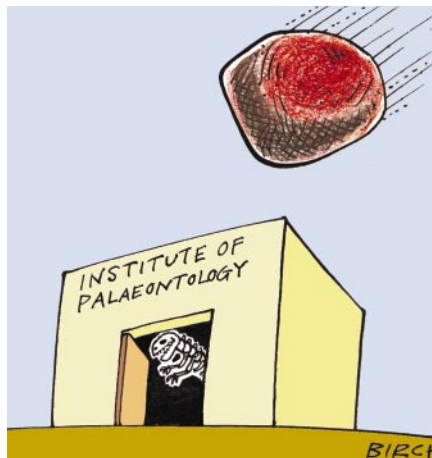
Hans-Peter Schultze, the museum's head of palaeontology, says that the proposed merger could mark the end of serious palaeontological research there, leaving behind only technicians to look after the fossil collection — one of the largest in Europe.

Schultze, who retires next year, also complains that there were no palaeontologists on the eight-strong panel. Alfred Crompton of the Museum of Comparative Zoology at Harvard University,

a palaeontologist who was supposed to represent the field, could not take part because no funds were available for him to travel to Berlin.

Gerhard Neuweiler, a zoologist at the University of Munich, former president of Germany's science council and a member of the panel, defends its findings. "We have not questioned the quality of palaeontological research at the museum," he says. Fears that it could be ended are unjustified, he adds.

But palaeontology is already in decline at several German institutes. The universities of Marburg, Darmstadt, Giessen and Mainz, for example, no longer offer courses in palaeontology or fill vacant positions. ■



Joint projects see ocean science aiming high

Mark Schroppe

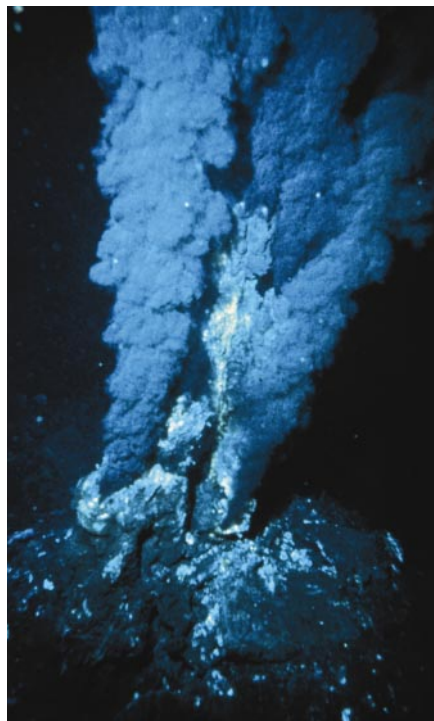
When former US President John F. Kennedy helped to launch the age of space exploration, he predicted that space would be the new ocean. But 40 years on, it is ocean scientists who are looking to space for the technologies and methodologies to fuel their own new age.

Despite some obvious crossovers between space and ocean sciences, the collaborations have never really met their full potential, says Craig McLean, director of the National Oceanic and Atmospheric Administration's Office of Ocean Exploration. But that is now starting to change.

Space scientists, for example, are aching to search for life in aquatic environments beyond Earth, such as Jupiter's moon Europa. Oceanographers can offer some of the tools for this research, and would in return benefit from space funding both to improve these instruments and to develop new equipment. And space scientists know that the oceans are an ideal testbed for tools, such as sensors, needed for their projects.

Last month at the Kennedy Space Center in Florida, McLean's office and NASA's oceanography programme co-sponsored a meeting called the Link Symposium, to foster such collaborations.

Many of the discussions at the meeting addressed shared needs for new technologies, such as better and smaller mass spectrometers that could serve a variety of purposes on both underwater vehicles and space probes.



Think big: those studying seafloor features are hoping to emulate the scope of NASA's projects.

Some such partnerships already exist. Ocean engineers, for example, have worked with their space counterparts to design vehicles to explore Europa's ocean. The two fields have also cooperated on the development and testing of sensor webs, networks of electronics pods that can be fitted with a variety

of sensors. "I think right now we are on the front end of this collaboration, which I am very confident will bear fruit," says McLean.

But the tendency of the two fields to think and operate on different scales presents a barrier to collaboration, says Eric Lindstrom, head of NASA's oceanography programme. "They don't speak the same language," he says. Space scientists tend to start with a long-term goal and then build what is needed, from scratch if necessary. Far smaller budgets have obliged oceanographers to strive for more incremental advances.

Some oceanographers hope to break this mould. For instance, Project Neptune is an ambitious plan led by researchers at the University of Washington in Seattle to wire an entire seafloor tectonic plate with sensors and equipment to monitor its geology, biology and chemistry.

At a cost of around \$250 million, the project, if it gets funded, would be comparable in scale to a typical NASA mission—making it the largest ocean-science project ever. Partly because of Neptune's similarities to space projects, NASA's Jet Propulsion Laboratory in Pasadena, California, is a partner in the project.

"Enormous changes have to go on in evolving from individual principal-investigator science to big science," says James Bellingham, director of engineering at the Monterey Bay Aquarium Research Institute in Moss Landing, California, and a coordinator of the Florida meeting.

♦ www.thelinkproject.org

US postal service puts anthrax detectors to the test

Jonathan Knight, San Francisco

Automated anthrax-sniffers will be making their debut in selected US post offices this summer. The US Postal Service (USPS) is already testing the equipment with an eye to installing it nationwide in all of its 400 or so mail-sorting facilities by next year.

The USPS hopes that the equipment, which detects the DNA of the anthrax bacterium, *Bacillus anthracis*, will prevent, or at least mitigate, the effects of any future bioterrorist attacks through the mail. In last October's attacks, anthrax spores contaminated several post offices after being dispersed into the air by mail-sorting machines, killing two postal workers.

A \$3.7-million trial to be carried out at several sorting offices on the east coast will test a detector produced by Northrop Grumman, an aerospace contractor based in Baltimore, Maryland. The device continuously channels samples into a

detection unit that screens for DNA from the anthrax bacterium.

A similar device designed by Lockheed Martin has not yet entered field tests. Whichever system is chosen will be installed nationwide and paid for, in part, by \$200 million in emergency funding that Congress set aside last autumn for the project.

The detectors in Northrop Grumman's system are made by the biotechnology company Cepheid of Sunnyvale, California. They rely on the polymerase chain reaction (PCR) to single out and amplify regions of DNA found in the toxicity gene of anthrax. A molecular tag turns fluorescent if it finds amplified DNA fragments to bind to, activating the alarm within half an hour of sample collection. If the gene is not present, no DNA is produced. A worker would periodically replenish the disposable reaction chambers in the detector, but the system requires little other maintenance.

All detection systems carry a risk of crying wolf with 'false-positive' results. Cepheid operations chief Kurt Petersen says that controls built into the PCR system should minimize these, but they won't know for sure how well they work until the tests are complete. For example, the system could run a second test to confirm a positive result before a facility is evacuated. "We are entering a realm of DNA testing that has never been done before," he says.

The trade-off in eliminating false positives is missing some genuine positives, says Martin Blaser, a microbiologist at the New York University Medical School who has studied last October's anthrax attacks. But he says it is better to risk missing some traces of anthrax than to panic postal workers and erode their trust in the system by raising too many false alarms. "We won't be able to develop a system that will protect everybody," he warns.

Australia lines up funds and feuds for new stem-cell centre

Sydney Stem-cell research continues to divide politicians around the world. But the Australian federal government is pressing ahead with plans for a new A\$43-million (US\$24-million) stem-cell research centre, just weeks before politicians begin debating legislation that could severely limit the research it can carry out.

The money will be used to establish the Centre for Stem Cells and Tissue Repair at Monash University, near Melbourne.

Researchers there will create new human embryonic and adult stem-cell lines, and attempt to develop cell-based treatments for diabetes, nervous-system disorders and other diseases.

Australian researchers are currently allowed to extract stem cells from surplus embryos made before 5 April this year, the date on which Australia's political leaders voted on the issue (see *Nature* **416**, 574; 2002). But Kevin Andrews, Australia's minister for ageing, who is responsible for stem-cell research, opposes research on human embryos, and parliament is expected to debate new stem-cell regulations next month.



Bleaching (right), whereby coral loses associated algae, affects over half of the Great Barrier Reef.

Rising temperatures blamed for coral's colour crisis

Sydney An aerial survey has confirmed fears that the Great Barrier Reef, the world's largest reef system, is suffering one of the worst coral 'bleaching' episodes on record.

Experts warned in February that this year's bleaching, which they attribute to mild increases in ocean temperature, could be the most severe in recent history (see *Nature* **415**, 947; 2002). Researchers have now assessed more than 640 reefs in the region, and found that almost 60% are affected. Extensive coral death was seen at some sites, but the scientists say they are hopeful that most of the reef will survive the bleaching episode relatively unscathed.

The survey was conducted by researchers

from the Australian Institute of Marine Science, the CRC Reef Research Centre and Great Barrier Reef Marine Park Authority, all based in Townsville, Queensland.

Japanese computers to network for success

Tokyo A shift in Japanese computing policy seems to be imminent after the Ministry of Education, Science, Sports and Culture unveiled plans to plough 70 billion yen (US\$565 million) into networked computing.

The money will be used to develop 'grid' technology — specialist hardware and software used to link high-performance computers. The plan signals a shift from the country's traditional policy of supporting individual supercomputers, such as its Earth Simulator, a climate- and geophysics-modelling system (see *Nature* **416**, 579–580; 2002).

"The time of the large battleship supercomputers such as the Earth Simulator is now over," says Satoshi Sekiguchi, a grid pioneer and director of the newly established Grid Technology Research Centre at the National Institute of Advanced Industrial Science and Technology in Tsukuba. "What Japan needs is a scientific computing and database infrastructure that reflects the needs of its scientists and engineers."

Oldest French journal says *bienvenu* to English papers

Paris The *langue de Molière* is set to play a more minor role in France's oldest scientific journal. English will now be the preferred language for articles published in *Comptes Rendus*, the proceedings of the French Academy of Sciences.

The academy says that the move is part of a modernizing drive designed to attract a wider international readership to its multi-disciplinary journal, which was founded in 1835. More editors and referees from outside France are also being appointed.

English usage in science, or any area of French public life, is limited by the government, which wants to preserve the country's national language. International scientific conferences organized by public research agencies, for example, must by law be in French or provide a simultaneous translation, and scientific publications must provide at least an abstract in French.

Satellite team baffled by senseless sensor

Tokyo An important sensor on Aqua, NASA's water-cycle-observation satellite, began malfunctioning last week, just three weeks after the satellite was launched.

Researchers are mystified as to why the AMSR-E unit, a microwave sensor that is supposed to collect data on factors such as sea surface temperature, has started to produce strings of zeros. AMSR-E, built by Japan's National Space Development Agency (NASDA), is important because it is the only sensor on Aqua that can collect data through heavy cloud cover. It is also the only device on board that can measure soil moisture levels.

Researchers at NASA and NASDA stress that it is not uncommon for such problems to arise, but say that they have no timetable for repairing the sensor. The fault with AMSR-E could delay this autumn's launch of ADEOS-2, a NASDA Earth-observation



Short shelf-life: the AMSR-E sensor began malfunctioning after just three weeks in orbit.

satellite that will carry a sensor with almost exactly the same design.

♦ <http://aqua.nasa.gov>

Scientists find BBC culinary drama hard to swallow

London In a unique fusion of technology and cutting-edge kitchen equipment that gives new meaning to the phrase 'domestic science', a new BBC drama about genetically modified plants suggests that these controversial crops can be cooked up using nothing more sophisticated than a food blender.

Researchers have lined up to point out scientific errors in *Fields of Gold*. One scene shows a rogue scientist using a blender to modify wheat with an antibiotic-resistance gene scavenged from hospital waste. The gene then jumps into bacteria, creating a superbug that infects people, birds and foxes.

Mark Tester, a plant scientist at the University of Cambridge who was consulted by the show's makers, says they ignored his advice. "The programme indicates that horizontal gene transfer between organisms is something that happens easily and readily," Tester says. "As I told them, this is not the case." The BBC says that the two-part series, to be screened on 8 and 9 June, is drama, and was never intended to be a documentary.

♦ www.bbc.co.uk/drama/fields_of_gold.shtml

Flying free

Pilotless aircraft could help monitor forest fires, or collect data over vast areas of ocean. But aviation authorities are reluctant to let them share the sky with other aircraft. Tom Clarke reports.

Veerabhadran Ramanathan has a dream: fleets of gossamer-winged, solar-powered robotic aircraft cruising in formation for tens of thousands of kilometres, high above the Indian Ocean. As they fly, they carry out the atmospheric experiments that he struggles to do now with planes and satellites. Planes cannot stay airborne for long enough, and satellites are too far above the Earth's surface. "We can't continue to rely on conventional aircraft," says Ramanathan, who is based at the Scripps Institution of Oceanography in La Jolla, California. "We need an equivalent of satellites in the atmosphere."

Ramanathan's dream could be realized with the help of unmanned aerial vehicles (UAVs) — aircraft that can fly without pilots or remote control. UAVs already exist, and are capable of doing the science he wants and much more. The US military's Predator UAV, for example, has seen service in Kosovo and Afghanistan. And as the cost of UAVs falls, the craft are ready to fly the hawk's nest and be put to peaceful use by scientists.

But getting UAVs airborne is proving tricky. Insurance costs are staggering and aviation authorities are unsure how to regulate the craft. Fears that they could be used by terrorists have grown after 11 September, making UAVs about as welcome in civilian airspace as a UFO. Before scientists can benefit from UAVs, the runway needs to be cleared of red tape.

UAVs have clear advantages over manned aircraft. For safety, take-off and landing are usually remote controlled. But once up, UAVs can stay aloft for days or weeks on end, following a predetermined path and guided by Global Positioning System satellites.

Some can fly higher than piloted craft. The altitude record for a powered aircraft, 29,500 metres, was set last year by Helios, a super-light, solar-powered UAV made by AeroVironment, an aeroengineering company based in Monrovia, California. More robust UAVs can cruise at heights considered too low for manned craft. And with no lives at risk, they can enter extreme environments such as thunderstorms or the smoke plumes above volcanoes.

Fire watchers

The vehicles have already been involved in some impressive experiments. Jim Brass, a remote-sensing expert at NASA's Ames Research Center in Moffett Field, California, has used Altus, a modified version of Predator, to study forest fires. His team equipped the aircraft with visual and infrared sensors that could see through smoke. In September 2001, they used these to follow the hot spots of a Californian wildfire, and had the images on the Internet in just 15 minutes.

But UAVs will really come into their own when they can be used to study interactions between Earth and atmosphere. Brass's team hopes eventually to fly a UAV over the Amazon. He thinks it could be used to track daily changes in the gases released by the forest. "We've never been able to run a 24-hour flight over an ecosystem before," he says.

Ramanathan wants similar capabilities. His project, the Indian Ocean Experiment, is designed to study the particles that seed cloud formation. The research area is the entire Indian

Ocean — far too large for conventional planes to cover.

So Ramanathan wants to use teams of three UAVs in parallel. One will monitor particle formation at the ocean's surface, another will fly through the clouds, measuring their chemical and physical properties, and a third will take readings from above the clouds. Plans are at an early stage, but Ramanathan has started talking to AeroVironment's Paul MacCready, the designer of Helios, about what kind of UAV might be suitable.

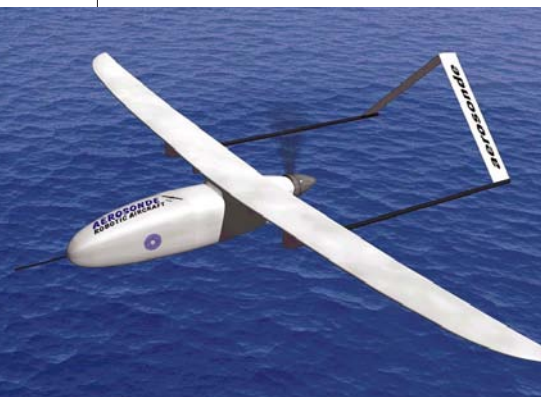
Aerosonde, a robotic-aircraft company based in Melbourne, Australia, is already providing its eponymous UAV to researchers. The aircraft is small, simple and efficient, and can carry enough equipment for atmospheric monitoring, weather prediction and some remote-sensing work, although not for the kind of project Ramanathan is contemplating. Last summer, Aerosondes participated in a NASA-run study of the hurricanes that batter the east coast of the United States during August and September. The UAVs were able to skim the ocean surface, measuring variables such as temperature and wind speed.

Other applications abound. Later this year, researchers at NASA's Environmental Research Aircraft and Sensor Technology (ERAST) project, based at the Dryden Flight Research Center in Edwards, California, plan to demonstrate the commercial potential of UAVs. They will use Pathfinder Plus, a high-altitude solar-powered craft similar to Helios, to monitor large coffee plantations, with the aim of enabling farmers to spot areas where the beans are ready to be picked. Other researchers want to use the vehicles for disaster monitoring or as high-flying relay stations for mobile-phone networks. When, in 1998, NASA asked scientists to apply for funding to develop UAV experiments, it received more than 60 applications.

But researchers' enthusiasm is not shared by everyone. Stephen Hottman, a UAV expert at New Mexico State University in Las Cruces, says insurance companies charge hundreds or thousands of dollars per hour to cover UAVs. And even if researchers can clear that hurdle, they can find the aviation authorities blocking their way.

There is no regulatory body for UAVs.

JON BECKER/AEROSONDE

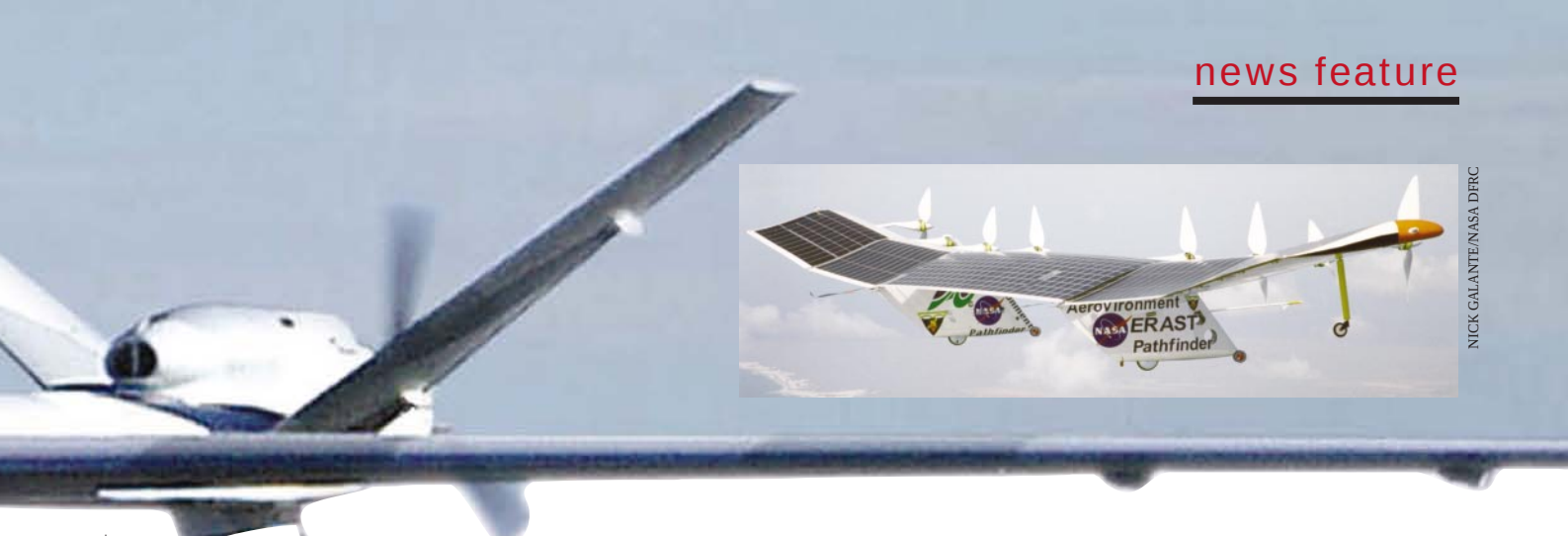


The pilotless Aerosonde has been used in hurricane research, and larger craft could help Veerabhadran Ramanathan study the Indian Ocean.



GENERAL ATOMICS AERONAUTICAL SYSTEMS

SIOTUCSD



NICK GALANTE/NASA DfRC

Eyes in the sky: unmanned craft have put forest fires on the Internet and Pathfinder (top right) will be checking the ripeness of coffee (right).



Division in Air Traffic, based in Washington DC, says the agency is "pressing on" with developing rules, and it hopes to incorporate UAVs into its regulations over the next 5 to 10 years. But with the exception of the military, requests for airspace from UAV users will still be dealt with on a case-by-case basis for now.

Although much of the investment in UAVs is in the United States, other countries seem to be ahead with their legislation. Australia already has rules for UAV flights, and Britain's Civil Aviation Authority is releasing a guide this month on how UAVs can be used and licensed to fly in UK airspace. It has also prepared draft rules on certification of UAV flightworthiness, design specifications and qualifications for operators. Other European countries are also drafting regulations.

US authorities might be encouraged to follow suit if new 'see-and-avoid' systems prove successful. These would ensure that UAVs don't fly into other aircraft or hit things on the ground. One system relies on aircraft and UAVs regularly sending out radio signals. If a craft picks up a signal, it sends out a warning in reply, alerting other aircraft and UAVs to its presence. The FAA plans to require all civilian aircraft to carry the technology needed to run this system, and it could be fitted to UAVs at relatively little cost.

Systems that do not depend on other aircraft could provide extra security. Aircraft radar has until recently been too large and power hungry to be carried by UAVs. But new lighter radar could be fitted to UAVs if the FAA required it. The Canadian company Amphitech of Laval in Quebec has



developed a radar system that scans a hemisphere-shaped volume of sky around 10 kilometres in radius, and is light enough to be fitted to the larger UAVs.

Hottman and colleagues are also championing the idea of a triangular flight corridor stretching from New Mexico to Alaska, across the Pacific to Hawaii and back to New Mexico, a route that includes airspace that is less busy than elsewhere in the United States. They hope to gain clearance from the FAA to fly regular UAV missions within this corridor. As other aircraft would have to cross it, Hottman hopes it would get the FAA used to UAVs sharing airspace, and help drive the development of rules for the operation of UAVs.

The FAA says the idea is under review. But if it does get approval, UAV researchers will finally have a test-track for their vehicles. And with the new see-and-avoid technologies coming up, aviation authorities may be more willing to let UAVs range farther afield. The demand from the scientific community is clearly there. If the FAA can be persuaded to come on board, disciplines ranging from atmospheric science to ecology could benefit from a new addition to their experimental toolbox. ■

Tom Clarke works in Nature's news syndication team.

Altus Fire Demo

♦ geo.arc.nasa.gov/sge/UAVFiRE

Indian Ocean Experiment

♦ www.indoex.ucsd.edu

Pathfinder coffee project

♦ www.clarku.edu/faculty/herwitz

DAVID MCNEW/GETTY

Researchers wanting to fly them in US airspace have to apply to the US Federal Aviation Administration (FAA) on a case-by-case basis. UAVs are always restricted to airspace where other craft rarely fly. Even so, each application can take weeks or months to process, and not all are successful. An application to fly an Altus over southern Florida, for example, was turned down because some of the flights, which were part of a plan to study electrical discharges from thunderstorms, crossed the airspace of the Kennedy Space Flight Center.

Is it a plane...?

Hottman, who is working on ways of integrating UAVs into airspace regulations, says that the FAA is still scratching its head over whether you can even register a UAV as a plane. The lack of guidelines creates a catch-22 situation. Until there are rules, there are no criteria for UAVs to meet. But because the FAA has little experience of dealing with UAVs, it is hard for it to develop rules. "It's chickens and eggs and donkeys and carrots," says Hottman.

But researchers on the ERAST project, together with others in the UAV industry, are now putting pressure on the FAA to get the ball rolling. New Vistas International, an aerospace consultancy based in Austin, Texas, recently submitted a white paper to the FAA outlining the case for accepting UAVs into the aviation industry. And Hottman's group has developed a 'roadmap' document that outlines how one type of UAV — high-altitude, long-endurance vehicles — might be licensed. This document, currently being considered by the FAA, suggests rules for everything from a UAV's headlights to the fuel it carries.

The FAA says it recognizes the need to license UAVs, if only to remain competitive with other countries. Alton Scott, manager of the administration's Special Operations



Cooking up a storm: Genentech's new antibody production plant in Vacaville, California.

Magic bullets hit the target

After decades of disappointment, antibodies are finally emerging as viable — if expensive — drugs. Trisha Gura finds biotech start-ups and pharmaceutical giants rushing to claim a piece of the action.

For more than a century, doctors have dreamt of using antibodies as 'magic bullets' to cure their patients. After all, these highly targeted proteins are among the immune system's key weapons. So if specific antibodies against proteins involved in disease could be produced in bulk, they should be ideal bespoke drugs.

The story so far has been a roller coaster of hope and disappointment. Despite advances in techniques for producing antibodies in the lab, the magic bullets have — until recently — performed poorly in the clinic. But thanks to developments in genetic engineering, cancer biology, immunology and genomics, antibodies might finally be on the brink of realizing their therapeutic potential. "Now everybody is going crazy to make monoclonals," says Leonard Presta, director of protein and antibody technology at DNAX, a biotech company in Palo Alto, California.

Monoclonals are antibodies mass-produced in the lab to recognize an individual molecular target. To date, the US Food and Drug Administration (FDA) has approved 11 of them — the majority in the past four years — to treat cancer and transplant rejection and to combat autoimmune diseases such as rheumatoid arthritis (see table, right). At least 400 other monoclonal antibodies are in clinical trials worldwide.

Antibodies are Y-shaped proteins, consisting of four polypeptides — two identical light and two identical heavy chains (see diagram, opposite). The arms of the Y identify and bind to the antibody's specific molecular target, or antigen. If that happens, the portions of the heavy chains that extend into the stem of the Y alert and recruit the other components of the immune system to attack the structure to which the antibody is bound.

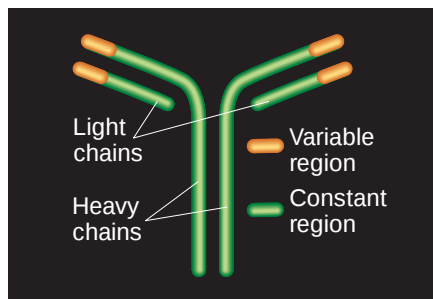
The therapeutic appeal of antibodies can be traced back more than a century, when mice were first investigated as a potential source. By injecting mice with infectious

agents, scientists aimed to stimulate the production of antibodies targeted against the infection. They hoped that they could then treat people suffering from the same condition by injecting them with the rodents' blood sera. But these crude preparations were ineffective, and the sera sparked adverse immune reactions in some unfortunate patients.

In 1975, however, Georges Köhler and César Milstein of the UK Medical Research Council's Laboratory of Molecular Biology (LMB) in Cambridge raised hopes with their invention of hybridoma technology¹, later honoured with a Nobel prize, which for the

Antibodies in action: monoclonal therapies approved by the US Food and Drug Administration

Product	Type	Condition	Company	Approved
Orthoclone OKT3	Mouse	Transplant rejection	Johnson & Johnson	June 1986
ReoPro	Chimaeric	Cardiovascular disease	Eli Lilly	December 1994
Rituxan	Chimaeric	Non-Hodgkin's lymphoma	Genentech	November 1997
Zenapax	Humanized	Transplant rejection	Roche	December 1997
Simulect	Chimaeric	Transplant rejection	Novartis	May 1998
Synagis	Humanized	Respiratory syncytial virus	Medimmune	June 1998
Remicade	Chimaeric	Rheumatoid arthritis/ Crohn's disease	Johnson & Johnson	August 1998
Herceptin	Humanized	Metastatic breast cancer	Genentech	September 1998
Mylotarg	Humanized (toxin-linked)	Acute myelogenous leukaemia	American Home Products	May 2000
Campath	Humanized	Chronic lymphocytic leukaemia	Millennium Pharmaceuticals	May 2001
Zevalin	Chimaeric (radionuclide-linked)	Non-Hodgkin's lymphoma	IDEC Pharmaceuticals	February 2002



Of mice and men: researchers have produced drugs based on (left, from top to bottom) mouse, chimaeric, humanized or human antibodies. All have the same basic structure (above).

grew into tumours that could produce large amounts of monoclonal antibodies. Today, improved cell-culture techniques mean that large quantities of monoclonals can be made without the need to grow tumours in mice.

Mouse antibodies worked well in rodent models of disease². But when doctors started injecting mouse monoclonals into human patients, problems emerged. The patients' immune systems quickly recognized the mouse antibodies as foreign proteins, and generated 'human anti-mouse antibodies' that cleared the mouse proteins from the bloodstream before they had chance to work. In rare cases, the result was a fatal allergic response².

Less is more

Researchers also soon realized that mouse antibodies could not function normally in people, because the structure of the proteins is subtly wrong. Even if the arms of a mouse monoclonal did manage to capture its corresponding antigen, the antibody's heavy chains could not signal the human immune system to attack the bound target³.

Indeed, only one mouse antibody made it through clinical trials. This monoclonal, called Orthoclone OKT3, was approved in 1986 and is used to help prevent the rejection of transplanted organs. It works by targeting a glycoprotein on the surface of T cells in the immune system that would otherwise recognize the organ as foreign. In effect, Orthoclone OKT3, marketed by Johnson & Johnson, shuts down one arm of the immune system — which helps to explain why it does not get cleared from the bloodstream quickly.

At the time, making completely human monoclonals was impracticable, because there was no human myeloma cell line suitable for making hybridomas. So scientists turned to genetic engineering of both mouse antibody-producing cells and hybridomas, mixing and matching DNA from mouse and human antibody genes in an attempt to make antibodies that would not be rejected by the human immune system. In the first of these chimaeras, genes encoding mouse antibody arms were simply grafted onto the genes for human antibody stems, producing antibodies that were roughly 30% mouse, 70% human.

These chimaeric antibodies could communicate with the human immune system⁴. But problems continued — in many cases, the chimaeras were still sufficiently 'mousey' to be attacked by the immune system⁵.

The breakthrough came in 1986, when Greg Winter's group at the LMB shaved the mouse component of chimaeric antibodies down to only 5–10% (ref. 6). Winter knew that three loops of amino acids within a part of each antibody arm called the variable region acted as the 'glue' to bind antibody to antigen. So he replaced everything but the genes for these loops with human sequences. The resulting 'humanized' mouse monoclonals seemed to evade the human immune system. "All of a sudden, it looked like maybe these things would become drugs," says Nils Lonberg, senior vice-president and scientific director of Medarex, a biotech company in Princeton, New Jersey.

As chimaeric-antibody technology and Winter's refined technique transferred to the biotech industry, a healthy pipeline of monoclonals lined up for FDA approval. In 1994, Eli Lilly of Indianapolis gained permission to market the first chimaeric antibody, ReoPro, which lessens the risk of blood clots in patients with cardiovascular disease by targeting a receptor protein on the surface of platelets. And in 1997, Roche of Basel, Switzerland, won approval for the first humanized monoclonal antibody, Zenapax — which combats organ rejection by binding to and inhibiting a receptor on activated white blood cells, which would otherwise stimulate tissue rejection.

The plethora of potential targets being identified by companies working in the fields of genomics and proteomics should fill the pipeline further. But so far, the main focus has been cancer, where a concentrated research effort has already identified many targets.

Solid start

After first going after leukaemia and lymphomas, with some success, companies have moved onto solid tumours, which are generally harder to treat. Herceptin, marketed by Genentech of South San Francisco, targets and blocks a growth receptor on the surface of breast cancer cells. In a trial of 469 women with late-stage breast cancer who tested positive for the receptor, a combination of Herceptin and standard chemotherapy caused greater tumour shrinkage and extended the patients' lives by an average of five months, compared with chemotherapy alone⁷. That, in the world of oncology, is considered a success. "Any incremental therapy, we embrace," says cancer researcher Rakesh Jain of the Massachusetts General Hospital in Boston.

The effectiveness of monoclonals against cancer might be improved by attaching toxins or radionuclides to the antibodies, to deliver a knockout punch to the targeted cells. In February, for instance, IDEC Pharmaceuticals of San Diego received

first time allowed researchers to mass-produce individual antibodies. By fusing antibody-producing cells from immunized mice with antibody-secreting mouse cells derived from a type of cancer called myeloma, they generated hybrid cell lines that could be cloned and cultured indefinitely. Injected into mice, these immortalized cells

approval for Zevalin, a radioantibody for use against non-Hodgkin's lymphoma. And some researchers are advocating the use of cocktails of monoclonal antibodies that will simultaneously target different molecules in multiple cell-signalling pathways. Jain has tried to do just that by testing Herceptin with another antibody, also made by Genentech, that inhibits the formation of blood vessels — needed to supply a growing tumour with oxygen and nutrients.

Intriguingly, Jain's team reported in March that Herceptin alone may do more than simply shut down the cancer cells' growth signal — it might also discourage the growth of blood vessels⁸. To some experts, findings such as these suggest that antibodies may have much broader applications than just inhibiting the function of particular proteins, or labelling them for attack by the human immune system. "You've got to think about antibodies as agents that change the biology of the cell," says Lee Nadler, senior vice-president of experimental medicine at the Dana-Farber Cancer Institute in Boston.

In that vein, Nadler wonders if antibodies might be used to help boost the effects of chemotherapy or to sensitize cells to radiation. Monoclonals are also being explored for their potential to bind to a receptor and activate it.

In parallel with the clinical developments, techniques for producing and selecting antibodies have also been advancing, allowing the generation of completely human monoclonals. A technology called phage display, for instance, allows researchers to build libraries of human antibody genes and incorporate them into bacteriophages, viruses that infect bacteria. The phages reproduce in cultures of *Escherichia coli*, and researchers can fish out any desired antibody from the resulting phage 'soup' using the appropriate target antigen tethered to a surface⁹.

Human behaviour

Meanwhile, Medarex and another biotech company, Abgenix of Fremont, California, are working with mice engineered to have immune systems that are human, as far as their production of antibodies is concerned. Scientists first knocked out the rodents' ability to produce mouse antibodies by deleting regions of the rodents' heavy- and light-chain genes, and then added the equivalent human genes^{10,11}. Now researchers just need to inject the rodents with the antigen of choice and the animals churn out completely human antibodies. "The beauty is that the mouse does everything for you," says Geoff Davis, chief scientific officer at Abgenix.

At the same time, researchers have not given up on making human hybridomas. Abraham Karpas of the University of Cambridge, UK, has spent years patiently cultivating a human myeloma cell line that can withstand fusion with other antibody-producing cells and still grow properly. He reported success last year¹². And in unpublished findings, Michael Neuberger at the LMB has come up with a way to speed up the generation of new antibodies using a line of antibody-producing cells that mutates its genes faster than normal.

But despite the current wave of enthusiasm, some problems remain. Investment in the antibody business recently took a hit as biotech company ImClone Systems of New York came under fire for inadequacies in its application to market a promising anti-tumour antibody (see 'Giving antibodies a bad name', left). But the biggest issue is cost. Although antibodies require much less investment in initial research and development than conventional small-molecule drugs, they are hugely expensive to manufacture.

This stems from the cost of scaling up antibody production to meet clinical demand. The proteins are usually generated by cell cultures in bioreactors that have a capacity of 40,000 litres or more. By the time the cells are nurtured, isolated and the antibodies they produce are purified, the cost can run as high as US\$1,000 per gram — compared with \$5 per gram for typical small molecules produced by chemical synthesis.

In addition, building a 100,000-litre facility, such as the one just completed by Genentech in Vacaville, California, can take five years and cost \$400 million. "There is this tension that as more and more antibodies come on line, there is a real problem with making enough," says Presta, who used to work for Genentech.

Right now, researchers are concentrating on trying to improve the efficiency of antibody production in cell culture. But they hope eventually to move to streamlined cell-free systems — a step that would surely place antibodies in the clinical mainstream. "We have a long way to go before the story is finished," says Winter.

Trisha Gura is a freelance writer in Cleveland, Ohio.

1. Köhler, G. & Milstein, C. *Nature* **256**, 495–497 (1975).
2. Winter, G. & Milstein, C. *Nature* **349**, 293–299 (1991).
3. Lubeck, M. D. et al. *J. Immunol.* **135**, 1299–1304 (1985).
4. Brüggemann, M. et al. *J. Exp. Med.* **166**, 1351–1361 (1987).
5. Brüggemann, M., Winter, G., Waldmann, H. & Neuberger, M. S. *J. Exp. Med.* **170**, 2153–2157 (1989).
6. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. *Nature* **321**, 522–525 (1986).
7. Slamon, D. J. et al. *N. Engl. J. Med.* **344**, 783–792 (2001).
8. Izumi, Y., Xu, L., di Tomaso, E., Fukumura, D. & Jain, R. K. *Nature* **416**, 279–280 (2002).
9. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. *Annu. Rev. Immunol.* **12**, 433–455 (1994).
10. Choi, T. K. et al. *Nature Genet.* **4**, 117–123 (1993).
11. Fishwild, D. M. et al. *Nature Biotechnol.* **14**, 845–851 (1996).
12. Karpas, A., Dremucheva, A. & Czepulkowski, B. H. *Proc. Natl Acad. Sci. USA* **98**, 1799–1804 (2001).

Giving antibodies a bad name

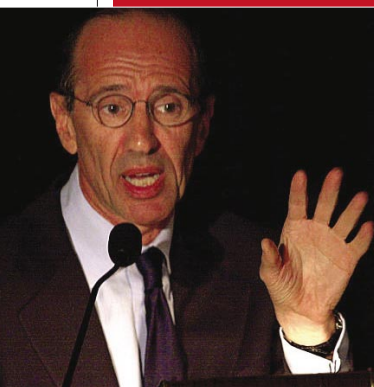
Encouraging results from clinical trials of anti-tumour monoclonal antibodies have done much to boost stock-market and scientific excitement. But the good work was nearly undone by the story of Erbitux, a drug developed by ImClone Systems of New York.

Erbitux is a monoclonal targeted against the epidermal growth factor receptor, found on about a third of solid tumours. Early studies of patients with colon cancer and tumours of the head and neck yielded such promising results that, in February last year, the US Food and Drug Administration (FDA) granted ImClone fast-track status for Erbitux. All that remained were further controlled, randomized trials. "We work very closely with the FDA," said Harlan Waksal, then chief operating officer of ImClone, last October. "They are our God."

But within two months, Waksal and his brother Samuel (below), then ImClone's chief executive, had provoked their deity's ire by rushing the trials. According to David Milroy, an analyst with the consultants Wood Mackenzie in Edinburgh, UK, the FDA first grew concerned about an apparent discrepancy that suggested colorectal patients on whom other treatments had failed were, in fact, responding to a standard chemotherapy agent. The agency asked for data confirming that the patients had indeed lost their ability to respond to chemotherapy. ImClone was unable to provide this, and could only supply complete records for three of 17 patients who had died of cancer. Researchers in the field add that the trials lacked proper randomization and appropriate controls. Indeed, experts accuse the Waksals of letting stunts such as hiring rock group the Doobie Brothers to perform at the American Society of Clinical Oncology meeting in San Francisco in May last year take priority over sound science. "ImClone is an embarrassment to those trying to do good research," says Lee Nadler, senior vice-president of experimental medicine at the Dana-Farber Cancer Institute in Boston.

For ImClone and others, the embarrassment has proved costly. The company is now facing investigation by the US Congress, the Department of Justice, and the Securities and Exchange Commission. Its stock fell from nearly \$74 per share in early December to a low of \$9.90 last month — and the ripples have spread. "Biotech in general and antibody companies in specific took a big hit," says Geoff Davis, chief scientific officer of Abgenix, an antibody company in Fremont, California.

The irony is that most industry sources believe Erbitux to be an effective drug. But rather than relying on ImClone's data, Merck KGaA of Darmstadt, Germany, which owns the rights to develop and market Erbitux in Europe, and ImClone's US partner Bristol-Myers Squibb are now running fresh clinical trials.



AP

Knowledge about animal suffering is too rarely used

Researchers in different fields should pool information to minimize unnecessary harm.

Sir — We at the Humane Society of the United States strongly agree with your Opinion article “Rights, wrongs and ignorance” (*Nature* **416**, 351; 2002), that the science of animal suffering and cognition should be given higher priority. The public is often told that the pain, distress and suffering of research animals are already being minimized. But given the lack of scientific knowledge about these issues, much more can and should be done to develop knowledge to identify, assess, avoid and/or alleviate such suffering.

Our current knowledge of animal suffering and cognition is relatively isolated, and the principles too rarely applied in animal research. Scientists engaged in research on farm-animal welfare, for example, should be consulted on how to investigate animal suffering in laboratories. Widely used techniques, such as the use of carbon dioxide for rodent euthanasia and the creation of genetically modified animals, should be reviewed and discussed, and a consensus should be reached on the level of suffering caused.

We challenge all stakeholders, including the research community, supervising agencies and the animal-protection community to tackle these issues together in a coordinated way (see www.hsus.org/ace/11350), to minimize animal suffering and to reduce the uneasiness that we all experience (biomedical scientists included) when animals are harmed in the cause of scientific progress.

In your earlier Opinion article “Time to cut regulations that protect only regulators” (*Nature* **414**, 379; 2001), you rightly state that the primary purpose of the Institutional Animal Care and Use Committees (IACUC) system should be the protection of laboratory animals from pain, distress and harm. We are no more in favour of unnecessary paperwork than the research community. We have proposed a revision of the pain and distress classification system that would focus attention on the riskier projects and could help avoid unnecessary paperwork for projects that carry no risk of pain and distress — but the research community has protested

vociferously against our suggestions.

You are, however, incorrect to state that IACUC reviews used to be carried out after a grant application had passed an initial approval stage. This has never been the case, although the system may indeed change soon, in line with the recent changes by the Institutional Review Boards (IRBs) for human subjects to the ‘just in time’ system you outline, to cut down unnecessary paperwork. You are also incorrect to state that cross-talk between institutions is not permitted: various publications advise on procedures in this regard, though we agree that far too little is done.

Any fear among institutions is being driven by the IRB system, not by the IACUC system. Death or injury to a human being generates risks of adverse publicity and litigation that are orders of magnitude greater than when laboratory animals are used.

Andrew N. Rowan

*The Humane Society of the United States,
2100 L St NW, Washington DC 20037, USA*

Conceptual biology: a semantic issue and more

Sir — Mikhail V. Blagosklonny and Arthur B. Pardee in their Concepts essay “Unearthing the gems” (*Nature* **416**, 373; 2002) discuss the emergence of ‘conceptual biology’ — the iterative process of analysing existing facts and models available in published literature to generate new hypotheses — and its value in integrating existing information conceptually as an essential part of scientific research. This process has parallels with the process of drug discovery by iteration of existing biomedical literature.

Scientists have traditionally worked in discrete communities, creating discipline-specific language. The natural consequence is that today we are faced with an overwhelming array of nomenclature for genes, proteins, drugs and even diseases.

The β -amyloid converting enzyme BACE, for example, has synonyms ASP2, memapsin 2 and β -secretase, because of different research groups working in different research communities. Catatonic schizophrenia is the more common term for what used to be known as schizophrenic catalepsy, schizophrenic flexibilitas cerea, catatonic dementia

praecox or schizophrenic catatonia. The drug pindolol has more than 30 synonyms, which include LB46, Carvisken, Calvisken and Prindolol.

In some areas, such as the naming of chemicals, there are controls that prevent names from being used more than once. In others, the same name has been used for different entities. The term ‘hunk’ is not only an English word, but also a cell type (human natural killer) and a gene (hormonally upregulated Neu-associated kinase). A search for the term ‘hunk’ in PubMed gives references to all three contexts.

The problem for scientists trying to perform ‘conceptual’ searches precisely and comprehensively is self-evident. It may be reasonably straightforward (though time-consuming) to search in a ‘concept-driven manner’ within their area of expertise, in which they are likely to be familiar with the language. It is considerably more difficult to search outside the discipline, as they may not know where to look and will be less familiar with variable nomenclature. As soon as a name is changed for a compound, disease or target, data associated with that entity could easily be lost. Delving back into history to seek those potentially valuable resources of information can often be impossible.

To address this all-too familiar problem, many groups are applying

standards to biomedical nomenclature; expert committees are reaching consensus on the naming of gene families, diseases and compounds; there are consortia such as Gene Ontology (www.geneontology.org), which is defining a controlled vocabulary for the association of genes across different species. From the opposite perspective, there are resources such as the Metathesaurus of the Unified Medical Language System (www.nlm.nih.gov/research/umls), which brings together numerous classification systems so that all alternative names and meanings defined by different groups are in one repository.

These are all extremely valuable resources, but what the biomedical community needs is the ability to use such resources to search the vast sources of information more effectively, to extract more meaning.

The annotation of Medline with medical subject headings (MeSH) is probably the best attempt so far (www.nlm.nih.gov/mesh), as it at least helps users to link their search-term to abstracts containing different terms with the same meaning. What is still needed is a way to control the context of the search, so that terms having different meaning in different contexts can be retrieved appropriately. We also need ways to enable scientists to cross disciplines and search in areas outside their expertise, so that they

can extract information critical for new discoveries. Knowledge-based systems will no doubt provide the best opportunity in this regard.

Julie C. Barnes

BioWisdom Limited, Babraham Hall, Babraham, Cambridge, CB2 4AT, UK

Uncovering the complex mysteries of mosaicism

Sir — I read with great interest your News feature “Dual identities”, bringing together intriguing results that are starting to emerge in a relatively unrecognized area of human genetics, chimaerism and mosaicism¹. It did not, however, cover single-gene mutations leading to somatic mosaicism, or germline mosaicism. I would like to draw readers’ attention to these developments because the clinical significance of these observations is probably more far-reaching than previously thought.

One of the initial lines of evidence implicating germline mosaicism in single-gene disorders was haplotype-analysis-based detection of the transmission of a muscular dystrophy gene by an unaffected male². Today, we know that a significant number of boys born with Duchenne muscular dystrophy are members of families in which one parent carries the disease-causing mutation only in his or her germ cells, an important issue in genetic counselling. It is increasingly being realized that mosaicism in germ cells can be an important underlying cause of disease in a growing number of genetic disorders³.

With respect to somatic mosaicism, we are now beginning to understand the biological mechanism by which some boys with X-linked dominant diseases such as Rett syndrome⁴ and incontinentia pigmenti⁵ survive.

Although it is not clear whether single-gene mutations mainly occur as a postzygotic event or before fertilization at the half-chromatid stage⁶, the presence of an X-linked dominant disease in a male is an important diagnostic sign of somatic mosaicism.

Tayfun Özcelik

Ayhan Sahenk Foundation and Department of Molecular Biology and Genetics, Bilkent University, Bilkent, Ankara 06533, Turkey

1. Pearson, H. *Nature* **417**, 10–11 (2002).

2. Darras, B. T. & Francke, U. *Nature* **329**, 556–558 (1987).

3. Zlotogora, J. *Hum. Genet.* **102**, 381–386 (1998).

4. Topçu, M. *et al. Eur. J. Hum. Genet.* **10**, 77–81 (2002).

5. The International Incontinentia Pigmenti Consortium. *Am. J. Hum. Genet.* **69**, 1210–1217 (2001).

6. Gartler, S. M. & Francke, U. *Am. J. Hum. Genet.* **27**, 218–223 (1975).

Bioinformatics code must enforce citation

Sir — Despite repeated calls for the development of open, interoperable databases and software systems in bioinformatics (for example refs 1–3), Lincoln Stein in his Commentary “Creating a bioinformatics nation”, with some justification compares the state of bioinformatics to the mediaeval city-states of Italy, and proposes a unifying code of conduct⁴. In considering his proposal, we must ask why such a chaotic situation arose, and why it has been so persistent.

There are many reasons for the existing chaos. Bioinformatics is a rapidly evolving field. Stable interfaces take time to design, implement and maintain. Algorithms and tools evolve and incorporate feedback from users, and the interfaces must necessarily evolve as well. But standards have been developed and widely accepted in other fields undergoing rapid technological change, such as the Internet.

The difference is that academic scientists are responsible for most of the software and data in bioinformatics. Academic careers are advanced by publications that establish priority and citations that validate the impact of the work. Being the first to develop a new approach forms the basis for a peer-reviewed publication, which is not the case for developing and maintaining a standard interface to an old tool or data set. Academic scientists cannot be expected to sacrifice their careers in the interest of community standards. Significant responsibility for standards development and implementation must fall to service organizations such as database providers. These organizations need the support of the academic community in standards development, whereas academic scientists need to benefit from the time and effort they contribute to the process.

Modern bioinformatics software systems are complex. A genome-annotation system, for example, draws on dozens of software components, involving teams of dozens or even hundreds of developers. We need ways to recognize the often-critical contributions of these individuals to the overall result. Stein’s code of conduct would facilitate the development of seamlessly interoperable systems in a way that hides the underlying complexity of a calculation from the user. From an academic scientist’s perspective, this goal is in direct conflict with the need for recognition and citation and will do nothing for the career of the developer. An academic scientist, therefore, has a strong career imperative to force users to deal

directly with their tool or website, and little incentive to make the technology accessible through interoperable systems.

BLAST⁵ and FASTA⁶ are “citation classics”; but they are also at the top of the list of “failure to be cited classics”. Projects like NCBI⁷ and Ensembl⁸ have made useful software tools and large volumes of data widely available, but do not give users the information necessary to cite appropriately the algorithms and software needed to access the system. Ensembl, for example, provides users with alignments performed by BLAST and SSAHA⁹ using EST sequences^{10,11} aligned to the human genome sequence¹² and gene models created by GenScan¹³. And yet Ensembl lists a citation only to itself on its home page, and the NCBI genome resources pages provide no citation information for the underlying bioinformatics.

If bioinformatics is to emerge as a strong “nation state”, Stein’s code of conduct needs to address the career imperatives of computational biologists. First and foremost, it must require people to cite their sources. Interfaces and data sets should include explicit citation information, so that systems assembled from components can recursively retrieve citation data from their components and present the user with information for all the modules used in a task.

Graphical user interfaces provide ready mechanisms to display the properties of an object. A user clicking on a gene model should be able to retrieve citation information quickly and automatically for the software and data used to assemble that model. This object- and task-specific citation approach would also provide a mechanism for recognizing the specific contributions of developers in large teams. The use of algorithms, software or data without attribution is plagiarism. Manuscripts that fail to cite bioinformatics sources properly are not acceptable for publication in peer-reviewed journals, and software systems that fail to cite their component sources are not appropriate for use by the scientific community.

David J. States

University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA

1. Boguski, M. S. *Curr. Opin. Genet. Dev.* **4**, 383–388 (1994).

2. Fasman, K. H. *J. Comput. Biol.* **1**, 165–171 (1994).

3. Karp, P. D. *Trends Biotechnol.* **14**, 273–279 (1996).

4. Stein, L. *Nature* **417**, 119–120 (2002).

5. Altschul, S. F. *et al. J. Mol. Biol.* **215**, 403–410 (1990).

6. Pearson, W. R. & Lipman, D. J. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).

7. Wheeler, D. L. *et al. Nucleic Acids Res.* **30**, 13–16 (2002).

8. Hubbard, T. *et al. Nucleic Acids Res.* **30**, 38–41 (2002).

9. Ning, Z., Cox, A. J. & Mullikin, J. C. *Genome Res.* **11**, 1725–1729 (2001).

10. Hillier, L. D. *et al. Genome Res.* **6**, 807–828 (1996).

11. Adams, M. D. *et al. Nature Genet.* **4**, 373–380 (1993).

12. International Human Genome Sequencing Consortium. *Nature* **409**, 860–921 (2001).

13. Burge, C. & Karlin, S. *J. Mol. Biol.* **268**, 78–94 (1997).

The times they are a-changin'

Policies on release of biological data should reflect reality, to the benefit of all.

Ari Patrinos and Dan Drell

Fierce controversy surrounds the issue of data sharing by authors of publications in major scientific journals. Private-sector authors are discouraged or shut out from publishing in the open literature because of demands for unlimited openness that are unacceptable to their corporate employers. We believe that ways should and can be found to maximize the amount of information that is openly shared.

In February last year, *Nature*¹ and *Science*² published the famous papers reporting the 'draft' sequence of the 3.2-billion-base-pair human genome. The *Nature* paper was by the publicly funded International Human Genome Sequencing Consortium; the *Science* paper was from Celera Genomics Corporation. Although the consortium's sequencing data were all rapidly released and deposited in GenBank as soon as available, Celera posted its data on its own website (www.celera.com) on publication, limiting free access to 1 million base pairs per day. Academic researchers requesting the entire sequence are required to sign a licensing agreement; those in the private sector have to negotiate a fee as a prerequisite for access.

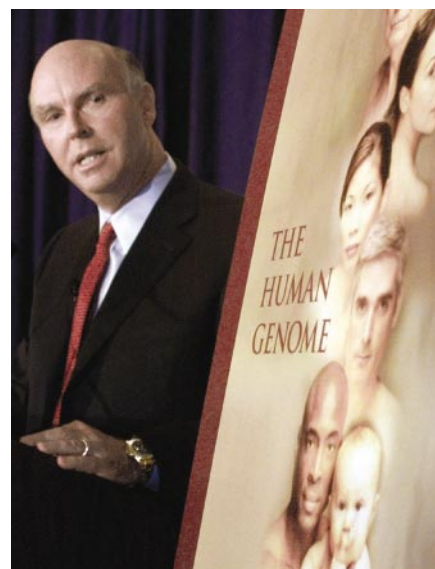
In April this year, *Science* published two papers^{3,4} reporting the draft genome sequences for two subspecies of rice, *Oryza sativa*. One, from Syngenta International, has limitations on data access essentially identical to those of Celera's human genome sequence. Much controversy has resulted from these constraints, with criticisms directed both at the private companies and at *Science*. Some critics believe that any limits on data access violate norms of standard scientific practice rooted in openness and unrestricted access to all data underlying a publication in the open literature. Eric Lander, first author of the human genome consortium's sequence paper, for example, believes that "if you choose to publish a claim, you must release all the 'integral data' supporting it, as determined by editors and peer reviewers"⁵. We have a different, more nuanced view, which we believe better acknowledges the realities of science as it is currently practised and funded.

Research funded by a private company is the intellectual property of that company, whereas data and other results derived from public (federal) funding are accessible not only to academic and other non-profit institutions but also to private companies. The challenge is to suggest how the private sector can be persuaded to share more data, to the benefit of all.

One possibility that has been proposed in



John Sulston (left) representing the public face of genomics and former Celera president Craig Venter.



various contexts^{6,7} is to start a timer on the deposition of certain data whereby a journal or other depository agrees to restrict access to the source data underlying a paper for a specified duration; or the data could be lodged with a trustee who ensures that the data were indeed deposited at the agreed time. Careful stipulations would be required both for how long the timer is set to run as well as precisely when it starts, but the idea is to permit a set duration for commercial exploitation (including filing of patent applications) on inventions derived from the data. The US Patent and Trademark Office allows up to one year before a provisional patent application is converted to a utility patent application, giving an applicant time to perform additional research towards clarifying the value of the invention while retaining the early priority date; thus one year might be a reasonable time for such a timer to run. This is similar to past practices at databases such as the Protein Data Bank.

The responsibility for implementing this scheme could rest with the journal or with a respected non-profit foundation (for example, the Institute for Scientific Information

or the Federation of American Societies for Experimental Biology). In consultation with GenBank (or a relevant public repository), the journal (or foundation) could provide access to the necessary files after the timer expires. It would be very useful to know the consequences of varying timer periods, as well as how much privately held data actually contribute to the commercial viability of a company, which could be investigated in a pilot study.

Careful timing

As time goes by, data lose value as new discoveries (and new technologies for reacquisition of the same data) are made, and as science inevitably proceeds unpredictably into new areas. The 'timer' mechanism would allow a company to publish valuable data that would otherwise remain private, while offering some protection for a limited duration for it to use the data exclusively. This role might be uncomfortable for journals and trustees, so it is important to explore fully a mechanism that all sides would have confidence in. An added concern could arise if implications for national or international security (for example, potential detection signatures in a pathogen's genomic sequence) emerged while the data were held on deposit before publication.

Another possibility would be to strengthen, perhaps even to codify in law, the research exemption by which a researcher using patented technologies is usually free of patent infringement fears if the research is fundamental and exploratory. This would not significantly affect the status quo, although

Is the academic scientific community willing to forgo the principle of universal free access?

its degree of protection would require careful legal definition and might be difficult to negotiate. However, such a law would permit the advancement of science using new technologies — such as PCR (polymerase chain reaction) or expressed sequence tag (EST) sequences as probes — without fear that the inventor of the technology would ‘reach through’ to claim intellectual property rights on new discoveries and thus discourage the original research. This protection would not extend to the use of someone else’s patented technology for an economic return, only for fundamental research intended for open publication and dissemination. The basic researcher would benefit from the use of new, powerful tools, and the entire scientific community, public and private, would benefit from the new knowledge gained.

Relaxing attitudes

There is an urgent need to find ways of giving incentives to the private sector, which now controls vast amounts of valuable data that have no obvious short-term commercial value but are of great potential research value. Most of the human genome sequence (about 98% of it) is non-coding; derestricting access to this part of it would not seem to threaten Celera’s stated goal of discovering candidate drugs based on those portions of the genome that encode expressed proteins. Additionally, the sheer volume of data from high-throughput sequencing centres challenges even the most advanced and sophisticated labs to mine it for value within a reasonable time frame.

Can incentives be defined to induce Celera, Syngenta and companies like them to relax access restrictions on this part of their data? These might take the form of bilateral exchange programmes through which academic scientists visit private-sector labs, contributing their knowledge and learning to a company’s science. University departments could offer opportunities, perhaps including adjunct appointments, to private-sector scientists to work in academic labs, providing legitimacy and academic status for work done by private researchers. A company could be enticed to collaborate by more direct funding of academic research that supports its ventures. For example, the Keck Graduate Institute in California has graduate students conducting company-sponsored research under confidentiality agreements in exchange for publication rights. Whatever the specifics, the benefits are in both directions: academic expertise and legitimacy would become more available to companies; private-sector research would become more accessible to academic scientists. Ideally, this becomes a ‘win-win’ situation for both sectors. Although such interactions have been common in other disciplines, the practice in biology is limited.

There are other precedents for successful

public–private collaboration. The SNP Consortium⁸ and the IMAGE Consortium⁹ both involved partnerships that placed into the public domain valuable genomic sequence information (the first on single-base-pair variants useful for trait mapping; the second on complementary DNA sequences representing expressed human genes). The commercial partners became valued contributors, having made the assessment that the value of restricting the data was less than the expected benefits of making them freely available. This is not without hazards, given the unpredictability of some commercial ventures (such as the corporate takeover and disbanding of Research Genetics, the distributor of one set of clone resources) and the long-term sustenance needs of databases. Although most research resources should be subject to the rules of free enterprise, there might be circumstances in which publicly funded agencies step in to fill gaps.

Whatever policy or policies are promulgated by scientific communities, it is the journals that, as a practical matter, must enforce them. The tradeoff has always been between the prestige of publishing in *Nature*, *Science* or other high-profile journals in exchange for openness and unrestricted access to the relevant data. What expectations of the journal review process are reasonable? In an era when the source data for a publication might be the complete multi-million base-pair sequence for an organism, which is obviously beyond the capacity of any journal to print, access must be via websites and the Internet.

A matter of trust

Can reviewers be temporary ‘trustees’ of the data under review, being permitted total access to the data in a paper submitted by commercial scientists to make a recommendation for publication subject to maintaining the confidentiality of the data? The notion of a trustee would require careful deliberation and a pilot experiment should explore its efficacy. It could help the scientific community to confront the key question of whether some constraints on data access are preferable to not seeing the data at all. We believe they are. Is the academic scientific community willing to forgo the principle of universal free access? If not, then an increasing fraction of the 60%



Rice research needs access to genome sequences

or so of genomics research that is conducted in the private sector will remain unavailable to academic and government scientists. In our view, that is too high a price to pay.

Our case is reinforced by the fact that most genomics companies do not publish their data. The many companies sequencing complementary DNAs and identifying single-nucleotide polymorphisms, for example, have information that would be immensely valuable to academic researchers if it were publicly available. But the companies patent genes as they are characterized and sell access to their databases under agreements that protect data as trade secrets. That is, of course, their right. But we need to create incentives for them to publish data within proprietary constraints, rather than clamouring for policies that push them towards non-disclosure.

The steady progress of science is founded on the tradition that individual scientists assemble knowledge ‘brick-by-brick’. We believe that full and unrestricted access to fundamental research data should remain a guiding star of science because centuries of experience suggest that it is the best way to achieve progress and realize science’s many benefits for everyone. However, we must also accept current realities.

Science has never been the exclusive province of the academic world. Yet today, the proportion of high-quality science in the private sector (the invention of PCR technology and the development of *cre-lox* recombinase gene knockout technology, to cite just two examples) is impressive as never before. The potential for productive collaboration has never been greater. We should not bemoan this development but welcome it. Private-sector science has its legitimate interests. The burden of argument is on the academic sector to attract and justify greater openness on the part of private-sector science by stating clearly what benefits this will bring to companies. Above all, it requires openness to new approaches, bereft of fundamentalism, regarding access to data that governments did not fund and cannot claim to own. ■

Ari Patrinos and Dan Drell are in the Office of Biological and Environmental Research, Office of Science, US Department of Energy, 19901 Germantown Road, Germantown, Maryland 20874-1290, USA.

1. The International Human Genome Sequencing Consortium. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* *Science* **291**, 1301–1304 (2001).
3. Yu, J. *et al.* *Science* **296**, 79–92 (2002).
4. Goff, S. A. *et al.* *Science* **296**, 92–100 (2002).
5. Marshall, E. *Science* **295**, 1625 (2002).
6. Petsko, G. *Genome Biol.* **3**, 1007.1–1007.2 (2002).
7. <http://beta.rcsb.org/pdb/nihi-policy.html>
8. <http://snp.cshl.org>
9. <http://image.llnl.gov>

Acknowledgements

We thank many colleagues for helpful comments, and in particular Robert Cook-Deegan for suggestions on earlier drafts. The views expressed here are those of the authors and do not reflect policy of either the US Department of Energy or the US Government.

Lateral thinking

A handy guide shows that there's more to left and right than meets the eye.

Right Hand, Left hand: The Origins of Asymmetry in Brains, Bodies, Atoms and Cultures

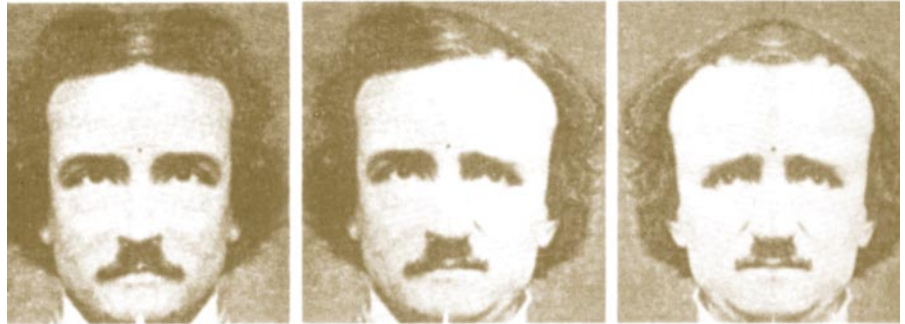
by Chris McManus

Weidenfeld & Nicolson: 2002. 256 pp. £20, \$27.95

William D. Hopkins

The causes and consequences associated with right- and left-handedness have been the focus of a growing body of research in the past 20 years. Indeed, laterality now even has its own journal: *Laterality: Asymmetries of Body, Brain and Cognition*, published by the Psychology Press. The scope and range of scientific disciplines now investigating laterality is the subject of this wonderful book by Chris McManus. Although its title implies that the focus is on handedness, don't be misled, for there is much more in this volume. The range of topics that it covers is far-reaching, and readers from a wide range of disciplines including physics, biology, chemistry, neuroscience and psychology will all find some aspects of the book intriguing from their own perspective.

Those who appreciate the history of science will not be disappointed either. McManus goes to great lengths to present a historical context for each of the topics discussed, and shows how the various disciplines are related and were often considered by our scientific founding fathers. He discusses topics ranging from asymmetry in the structure of amino acids to the socio-linguistic origins of the words 'left' and 'right'. He treats readers to a diverse set of observations, such as the fact that the skin of certain species of toad is extremely psychoactive because of differences in the distributions of left- and right-handed



Taking sides: composites of the halves of Edgar Allan Poe's face (centre) highlight human asymmetry.

amino acids. And there is a wonderful discussion of the religious and political meanings of 'right' and 'left' in relation to the origins of such words as 'sinistrality' or the common reference to the left hand as the 'sinister' one.

McManus begins by describing a patient who exhibited situs inversus, a medical condition in which an individual has the organs reversed, with, for example, the heart on the right side instead of the left. The patient is right-handed, a point that serves as the juxtaposition for much of the book. Most people are right-handed and most people have their heart on the left side of the chest cavity. Therefore, because individuals with situs inversus have reversed organ asymmetries, one might assume that they would be left-handed. Modern research has shown that this assumption is wrong, but the example is revealing because it extends the concept of laterality beyond the traditional realm of handedness. The right-handed situs inversus patient represents the dual definition of laterality, defined first as a morphological phenomenon and secondly as a behavioural and functional process. Each

form of laterality is autonomous from the other, but both are equally puzzling in terms of their origins.

From this basic premise, McManus begins by describing the historical and philosophical questions of why asymmetry exists. In particular he discusses the deeper question of right and left as absolute or relative degrees of space. Right and left are relative to the perspective of the viewer, so one need look no further than into a mirror to understand the fundamental problem.

In his discussion of the origins of human handedness, McManus presents a fair and representative review of various genetic, developmental and cultural models of handedness that have arisen over the years. He uses his own genetic model to explain individual variation in the expression of right- and left-handedness as well as variation between different cultures. Finally, he discusses several myths, misunderstandings and idiosyncrasies of left-handedness. For example, it is noted that a significantly higher proportion of left-handed individuals have been president of the United States compared with the population as a whole, and that left-handed individuals do not really have a shorter lifespan than right-handed individuals, as some believe.

This part of the book is a sometimes amusing, light-hearted addition to the otherwise empirically driven arguments in the preceding chapters, and many readers will find them both delightful and informative. I was disappointed that greater attention was not given to the plethora of recent findings on limb preferences in animals, including lower and higher vertebrates. Although some of these findings are referenced and briefly discussed, they are not fully presented — a shame in a book that purports to describe the origins of handedness. However, several recent books have been dedicated to the topic of behavioural laterality in animals, such as *Comparative Vertebrate Lateralization* (Cambridge University Press,



2002) edited by Lesley Rogers and Richard Andrews. Perhaps McManus opted not to include a lengthy discussion in his book to save space and to avoid distracting readers from the larger issues at hand.

Minor criticisms aside, he has done a marvellous job of summarizing and integrating a wide range of findings from various disciplines addressing questions on the nature of right and left. The presentation and clarity of the topics is palatable to both the scientific and lay communities, making this volume well worth reading. ■

William D. Hopkins is in the Division of Psychobiology, Yerkes National Primate Research Center, 954 Gatewood Road, Atlanta, Georgia 30322, USA.

Controlling a generous host

Parasites and the Behavior of Animals

by Janice Moore

Oxford University Press: 2002. 338 pp. £65, \$85 (hbk); £24.95, \$45 (pbk)

Paul Schmid-Hempel

In the early 1980s, a paper in *Scientific American* explained that a pill-bug (or woodlouse) is not always a pill-bug because its behaviour may be dramatically changed by an internal parasite to suit the latter's selfish interests. Although this was not the first summary of such parasite-induced alterations of host behaviour, it arrived in a climate of a suddenly increased awareness of the role of parasites in host ecology and evolution. Since then, the field as a whole has witnessed an unprecedented increase in the number of studies and publications.

Contrary to earlier beliefs, research over the past two decades has shown that parasites are not the exotic case nor a dull, degenerate lot that seek a harmonious coexistence with their hosts. Rather, parasites are ubiquitous and are continuously, relentlessly coevolving to overcome their hosts' defence systems. Parasite strategies, too, are governed by short-term fitness maximization even if it means that the host is not always killed or made sick. In doing so, parasites face the central dilemma of inflicting damage to the current host while still needing transmission to the next — a problem elegantly captured in Roy Anderson and Robert May's classical equation for R_0 , the parasite's net reproductive rate.

Even though we are still far from a comprehensive understanding, the application of such population and evolutionary thinking to host-parasite interactions is a success story. Strangely, as the field progressed, investigations into how and why parasites



Tripping the light fantastic

In the 40-odd years since Yuri Gagarin was the first man in space, some 400 others — including Linda Godwin (left) and Jerry Ross, shown above in the weightless conditions on board Atlantis —

have made the same remarkable journey. Serge Brunier looks back at the highs and lows in *Space Odyssey: The First Forty Years of Space Exploration* (Cambridge University Press, £25).

affect host behaviour fell behind. Janice Moore, the author of the *Scientific American* paper on pill-bugs, now sets out to correct this imbalance.

Her book is a gripping account of the sometimes spectacular behavioural (and morphological) alterations caused by parasites. She concentrates on infections by helminths ('worms'), such as nematodes, cestodes and trematodes, but regularly discusses fungi and protozoa too. Did you know that the first experimental demonstration of how tapeworms infect humans was made with the help of the guillotine? (See p. 20 of the book if you want the gruesome details).

Moore's writing is witty and conveys the flavour of her deep interest in beasts that most people like to avoid. Her organizing principle, which unfortunately is not carried through consistently, is the equation for R_0 . In fact, altered host behaviour directly affects most of the equation's variables. For example, many parasites render the host fearless. This increases the probability that the current host will be eaten by a predator that may then serve as the next host. Conversely, host avoidance behaviour reduces the risk of infection. Both processes change the numerical value of the transmission rate in the basic equation. Moore does not restrict herself to simple behaviours, however, but sensibly includes changes in life-history parameters, too. For example,

behavioural fever and behavioural chill, in response to infection, each reduce mortality rates and may accelerate recovery.

For the population biologist, behavioural alterations count if the parameters of the equations vary. Behaviourists like to trust their interpretation of the meaning of an observed behaviour. Physiologists want to know how parasites achieve their manipulative tricks. Moore's treatise gives all sides something to chew on, but it does not generate many new concepts to fill the gaps.

To pick a few faults, figures are not always attractive or well explained, quotations are occasionally unfortunate, and there are some unnecessary repetitions. But these minor shortcomings aside, the book is a pleasant read (even though some examples can send chills up your spine), a highly stimulating survey of the field, and certainly a reference for years to come. Everyone with a general interest in biology who can still be amazed by the awesome power of natural selection should look at this book. Along the way, you will learn many things and perhaps come up with a question that might change your own research. In this sense, the book is an important step. Hopefully, the next such step is not another 20 years away. ■

Paul Schmid-Hempel is in the Department of Environmental Sciences, ETH Zurich, CH-8092 Zurich, Switzerland.

Exploiting a hostile world

Life at the Limits: Organisms in Extreme Environments

by David A. Wharton

Cambridge University Press: 2002. 320 pp.

£18.95, \$25

Lynn J. Rothschild

It is a lovely day in the San Francisco Bay area. But then, I'm a *Homo sapiens* with a rather parochial definition of 'lovely' — roughly 20–27 °C, mostly sunny, atmospheric pressure of 1 bar, 21% O₂ and so on. "Frigid!" cries the thermophile. "I can't take the low pressure," declares the barophile. "Too much oxygen," gasps the anaerobe. Whole biotas exist in the environmental extremes, having already boldly gone where no human could.

Life in extreme environments is, forgive the pun, a very 'hot' subject. The idea that life can live at physical and chemical extremes thought previously to be well beyond the capabilities of life has enormous implications for fields as disparate as biodiversity and biotechnology. And what could be more exciting than the possibility that we Earthlings — from archaeans to anthropoids — are not alone in the vastness of space? The environmental space of life on Earth defines for us the minimum limits for life. From what we now know, it's not the sky that's the limit, but the availability of liquid water.

Life at the Limits is written with a disregard of current dogma. Life in extreme environments is far more than just which Archaea can stand the hottest bath, although the thought of *Pyrolobus fumarii* happily metabolizing at 113 °C is astounding.

The book begins with a philosophical discussion of what is extreme, and comes to the conclusion that there are certain chemical and physical extremes that should make life increasingly difficult. I agree, but differ from Wharton in how oxygen is treated. Certainly



Extremely difficult: heat and cold pose challenges to survival, but life seems to have the problem licked.

an aerobe that finds itself in a low-oxygen environment is stressed (if not dead), but Rocco Mancinelli and I argued in a review (*Nature* **409**, 1092–1101; 2001) that the ability to be an aerobe is a response to an extreme stressor: oxygen, along with its reduced by-products such as the hydroxyl radical and superoxide anion. So even humans can be considered as extremophiles because of their superior ability to thrive in an atmosphere containing as much as 21% oxygen.

The great strength of Wharton's book lies in his expertise in zoology. Most workers on extremophiles focus on microbes, often just on Archaea. But this bias is unjustified, as Wharton clearly demonstrates here. One of my favourite parts is on life in deserts, where desert vertebrates (yes, including the

camel), invertebrates and plants are given more attention than the microbes.

But perhaps this bias has gone a bit far for a balanced view of life in extreme environments. In some chapters a brief discussion of microbes appears almost as an afterthought. And I would be stripped of my space-microbiology credentials if I didn't question Wharton's uncritical acceptance of the Apollo 12 myth. In 1969 the Apollo 12 astronauts recovered the Surveyor 3 camera body, which had been on the moon since 1967. The human respiratory bacterium *Streptococcus mitis* was found in the camera body. However, after double-checking with NASA's planetary protection officer, John Rummel, I was able to confirm that this story is almost certainly incorrect — the source of the bacterium was probably a breach in sterile technique during the sampling of the camera. But the discussion of microbes in ice nucleation, and even in snow and ice-cream production, was fascinating. And there are ample alternative sources of information on extremophile microbes.

Would I have purchased this book? Probably not, because I was smug enough to think that I already had an excellent grasp of the field. But that would have been my loss, and I know that my students and I will enjoy consulting this book in the years to come for its coverage, enjoyable style and background material. ■

Lynn Rothschild is at NASA Ames Research Center, Moffett Field, California 94303-1000, USA.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the 7 November issue. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 2000 and published at least four separate issues by the end of June 2002.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language must be English.
- Deadline for submissions is 15 July 2002.

For each eligible title, please send at least four different issues (the first, the most recent and any two others), together with full details of subscription rates, to Mary Purton, *Nature*, The Macmillan Building, 4 Crinan Street, London N1 9XW, UK. Tel: +44 (0)20 7843 4567. Fax: +44 (0)20 7843 4596/4763. E-mail: m.purton@nature.com

Fibonacci's flowers

Amar J. S. Klar

The spiral arrangements of leaves on a stem, and the number of petals, sepals and spirals in flower heads during the development of most plants, represent successive numbers in the famous series discovered in the thirteenth century by the Italian mathematician Fibonacci, in which each number is the sum of the previous two (1, 1, 2, 3, 5, 8, 13, 21, 34, 55...). Seeds on the heads of sunflowers, for example, are arranged in two sets of spiral rows, one curving to the left and the other to the right. Thus, if 34 seed rows curve clockwise, there will be either 21 or 55 anticlockwise spirals on a sunflower head. Pine cones (see picture) are found both in the 'dexter' (righty) form, in which most spirals run clockwise, and in the 'sinister' (lefty) form, in which anticlockwise spirals predominate.

This set of phenomena is called phyllotaxis, from the Greek (*phyllon*—leaf; *taxis*—order). Phyllotactic patterns have been described for centuries, but the mechanisms that initiate these patterns remain undefined. The geometric arrangement follows from the regular packing of leaf primordia on a stem as the diameter of the stem slowly increases. But how does this pattern of growth conform to the numbers in Fibonacci's series?

The meristematic tissue, an undifferentiated mass of cells at the tip of a plant shoot, has near its boundary a region called the apical ring, where new plant organs are formed through extensive cell division in a structure called the primordium. Thus, phyllotactic patterns are thought to result from regulated differentiation of the primordia from cells originally derived from the meristematic

tissue of the vegetative or floral shoot. So in flower development, it seems that the floral tip produces seeds in spiral arrays as a result of spiral growth combined with primordia moving radially away from the centre of the apex.

Two main hypotheses have been proposed to explain the generation and maintenance of phyllotactic patterns. In the famous 'field' model, the position of the primordium is determined by undulating inhibitory fields, presumably composed of biochemicals, that emanate from the existing primordium and the apical meristem. The second model suggests that tissue mechanics and biophysical forces combine to promote morphogenesis in predictable ways. Yet neither of these models produces testable predictions, and they both lack convincing experimental support. Furthermore, they explain only the propagation of established patterns, not how phyllotaxy actually originates.

The connection between mathematical number series and pattern development remains to be described in biological terms. I would like to propose another, simpler theoretical model, based on cellular differentiation, to explain the *de novo* generation of phyllotaxy. Imagine an asymmetric cell division that gives rise to a mature cell that is competent to divide, as well as a juvenile cell that must first grow for one more length of the cell cycle to mature before it begins its division cycle. Remarkably, such an asymmetric cell division will indeed produce cell numbers in each generation that match the Fibonacci series.

This outcome is analogous to the original mathematical challenge posed by Fibonacci for his high-school students, to calculate the numbers of breeding rabbits when the

Plant mathematics

Asymmetric cell division offers a possible explanation of the spiral patterns seen in many plants.

newborns have to grow before they can begin breeding. Another striking analogy concerns stem cells—undifferentiated cells that divide to renew themselves as well as to give rise to more specialized cell types. Many cases exist in biology in which one daughter cell maintains the stem-cell characteristic while the other daughter is differentiated. For example, in early divisions of embryos of the nematode *Caenorhabditis elegans*, the times taken by different daughter cells to divide are very different. I suspect that a similar cell-division pattern may underlie the development of mathematical patterns in plants.

Intuitively, the stem-cell proposal predicts that floral meristems growing spirally and dividing asymmetrically will produce dexter and sinister arrangements in equal proportion, an outcome that is not predicted in such a straightforward way by the other models discussed above. Of 37 cones picked from a pine tree, I found that 20 cones were dexter and 17 were sinister, which is consistent with the idea that the direction of asymmetry is random. Likewise, six other trees all produced both kinds of cone. Randomness is expected in binary systems in which no bias exists, such as the tossing of a coin or the development of a crusher or clipper claw on the left or right side in lobsters. The currently prevailing models are unsatisfactory for answering such a fundamental question in biology.

At the very least, the stem-cell model is attractive in its simplicity compared with other models, and may provide a new framework for explaining existing results as well as becoming a concept that will guide research. At present, the challenge is to correlate asymmetric patterns of cell division with the generation of Fibonacci patterns, and to design tests to distinguish between these models. Perhaps the most useful approach may be to study mutants with altered developmental patterns. ■

Amar J. S. Klar is at the NCI Center for Cancer Research, National Institutes of Health, Frederick, Maryland 21702-1201, USA.

FURTHER READING

Sussex, I. M. *Cell* **56**, 225–229 (1989).
Turing, A. M. *Phil. Trans. R. Soc. Lond. B* **237**, 37–72 (1951).
Green, P. B. *Am. Zool.* **27**, 657–673 (1987).
Klar, A. J. S. *EMBO J.* **9**, 1407–1415 (1990).



Something sinister: the pine cone on the left is in the 'lefty' form; that on the right is dexter, or 'righty'.

Iodine's air of importance

Charles E. Kolb

Iodine-containing emissions from marine algae can be converted by sunlight into aerosol particles. If this phenomenon occurs on a large scale, it could have significant effects on climate.

The discovery of a previously unrecognized source of aerosol particles is big news to atmospheric scientists. Just such a source is described by O'Dowd *et al.* on page 632 of this issue¹. Building on earlier work, they have unravelled a photochemical phenomenon that occurs in sea air and produces aerosol particles composed largely of iodine oxides. The precursor molecules are organic iodide vapours emitted by marine algae.

One reason for the interest in atmospheric aerosols is their effect on climate and on our understanding of climate change². In particular, uncertainties about the composition and distribution of fine aerosol particles, no more than a few micrometres in diameter, cause large uncertainties in predictions of global warming driven by the accumulation of greenhouse gases. Depending on their composition, aerosols can have a direct effect on Earth's radiative balance by back-scattering (or absorbing) incoming solar radiation, leading to cooling (or warming). Indirectly, their influence is felt through their action as cloud-condensation nuclei, catalysing cloud formation. The more aerosol particles that can induce droplet formation in a cooling air mass, the smaller the resulting cloud droplets. Smaller droplets produce brighter clouds, which might also be longer lived because they are less likely to precipitate as rainfall. Indeed certain observations indicate that aerosols from forest fires and urban pollution can suppress rain and snow fall^{3,4}.

Aerosol sources are shown in Fig. 1 and can be divided into two types. Primary aerosols are emitted directly, such as smoke from bush or forest fires, soot and ash from factories, motor vehicles, trains, boats and planes, airborne dust, and sea-salt particles produced when sea spray dries out. Globally, however, much of the ambient particulate burden, and most of the fine aerosols, are produced in the atmosphere itself. These secondary aerosols arise from oxidation of precursor gases, such as sulphur dioxide, nitrogen oxides and volatile organic compounds, to form less volatile products. The resulting oxidation products then nucleate to form new particles or condense on pre-existing particles. Figure 1 also indicates the main effects of atmospheric particles.

For many years it was assumed that the primary chemical source of new atmospheric particles was the co-condensation of

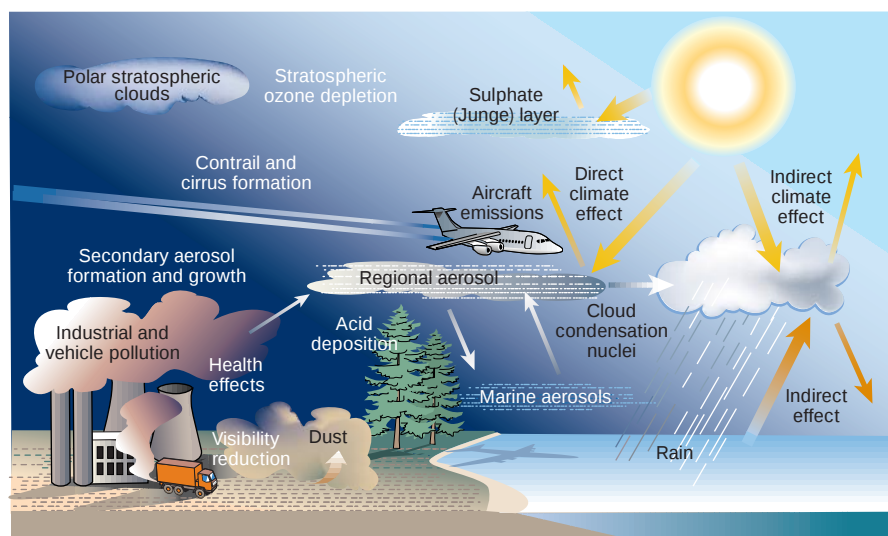


Figure 1 Aerosols — the big picture. Industrial and vehicle exhaust emissions, windblown dust, and salt from dried sea spray are all sources of primary aerosol particles. Secondary aerosol particles are produced in the atmosphere from gaseous pollutants in exhaust emissions, and emissions from land vegetation and marine organisms. Photochemical processes in urban smog produce high levels of secondary particles; lower, but still significant, levels stem from similar processes higher in the atmosphere. Atmospheric particles have many effects. Reactions catalysed in polar stratospheric clouds, and in the lower-latitude stratospheric sulphate (Junge) layer, result in ozone depletion. Photochemically produced particles of sulphuric acid, and nitric acid, lead to acid deposition. Fine aerosol particles of both primary and secondary origins can affect human health, reduce visibility, and influence climate both directly and indirectly. Particles and precursor gases emitted by aircraft in the upper troposphere and stratosphere can have a disproportionate effect because these regions are not heavily polluted by ground-level emissions. The mechanism discussed by O'Dowd *et al.*¹ might be an important contributor to the marine aerosol layer, and especially the tropospheric layer immediately above: sea-salt aerosols are fairly large and are generally not transported far above the surface.

sulphuric acid vapour and water vapour⁵. Over continental areas, sulphuric acid vapour is formed primarily by the oxidation of sulphur dioxide produced in burning oil and coal. In clean marine environments it is produced by the oxidation of dimethylsulphide and other reduced sulphur compounds emitted by marine organisms⁶.

With improved measurement techniques, however, bursts of formation of new particles have been observed when the concentration of sulphuric acid vapour is too low to support its combination with water vapour. In some special cases, atmospheric concentrations of other condensable inorganic species, for instance ammonia or nitric acid vapour, are high enough to account for particle formation and growth through a nucleation mechanism⁷. And quite recently, evidence has emerged for a completely different source in forested regions: gaseous monoterpenes released from trees can be

photo-oxidized to condensable carboxylic acids fast enough to produce bursts of particle formation and growth^{8,9}.

Over the past few years, O'Dowd and co-workers have observed episodes of particle creation in their study areas along the Irish coast. But these episodes could not be explained by nucleation and condensation driven by sulphuric acid or carboxylic acids^{10–12}. Knowing that seaweed can emit easily photolysed alkyl iodide compounds, such as CH_2I_2 , and that the resulting gaseous iodine reacts rapidly with ozone and other atmospheric oxidants to produce iodine oxides, they explored this avenue through laboratory experiments. This preliminary research showed that the photolysis of CH_2I_2 in the presence of ozone produces copious numbers of fine particles¹³.

In their latest paper¹, O'Dowd and colleagues have taken the earlier work a crucial step further. They have extended the

laboratory experiments to realistic coastal conditions, reproduced in a state-of-the-art atmospheric smog chamber. They show that the photolysis of CH_2I_2 at concentrations as low as 0.015 parts per billion by volume, well within levels often found in coastal environments, is a potent source of aerosol particles. Using a suite of instruments for characterizing the dynamics of particle formation and growth, and an aerosol mass spectrometer for determining chemical content as a function of particle size, they charted particle dynamics and confirmed that the particles formed in their chamber were predominantly composed of iodine oxides, the simplest of which may be OIO , HOI and I_2O_2 . The authors' suggested reaction mechanism for the creation of these species, after the photochemical production of iodine from algal CH_2I_2 emissions, is shown in their Fig. 2 on page 633.

To produce stable new particles in the clean, open-ocean marine atmosphere, concentrations of condensable vapour have to be high enough both to nucleate new nanometre-scale particles and to allow them to grow by agglomeration and vapour condensation to the stable 50–100-nm size range⁵. If there is too little condensable vapour, new particles don't form or they re-evaporate or agglomerate with pre-existing particles. O'Dowd *et al.*¹ describe modelling calculations which suggest that CH_2I_2 concentrations over the open ocean might well be high enough for the resulting condensable iodine oxides to allow newly nucleated sulphuric acid particles to become large enough to survive. The authors propose that the resulting particles might be abundant enough to influence the Earth's radiative balance. At the least, their model suggests that iodine oxides produced from volatile organic iodide compounds such as CH_2I_2 must be added to the list of precursors for secondary aerosol formation.

In retrospect, this might not be too surprising. In pioneering research off Hawaii¹⁴ and Puerto Rico¹⁵ in the 1970s, it was shown that iodine becomes concentrated in atmospherically processed sea-salt aerosol. In contrast, other halogens—chlorine and bromine—are depleted. These 30-year-old studies further showed that the iodine levels vary inversely with particle size, just as one would expect from a gas-phase condensable source of iodine oxide such as that described by O'Dowd *et al.*¹

The obvious task that remains is to determine just how widespread this newly identified mechanism of particle growth is. To have a significant influence on climate, it would have to be effective over the oceans as a whole, not just in the coastal environment. The appropriate field-measurement tools and analytical models are already in hand. ■

Charles E. Kolb is at Aerodyne Research, Inc., 45 Manning Road, Billerica, Massachusetts 01821-3976, USA.

e-mail: kolb@aerodyne.com

1. O'Dowd, C. D. *et al.* *Nature* **417**, 632–636 (2002).
2. Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, 2001).
3. Rosenfeld, D. *Geophys. Res. Lett.* **26**, 3105–3108 (1999).
4. Rosenfeld, D. *Science* **287**, 1793–1796 (2000).
5. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics* (Wiley, New York, 1998).
6. Andreae, M. O. & Crutzen, P. J. *Science* **276**, 1052–1058 (1997).
7. Weber, R. J. *et al.* *Geophys. Res. Lett.* **26**, 307–310 (1999).

8. Kavouras, I. G. *et al.* *Nature* **395**, 683–686 (1998).
9. O'Dowd, C. D. *et al.* *Nature* **416**, 497–498 (2002).
10. O'Dowd, C. D. *et al.* *Geophys. Res. Lett.* **25**, 1661–1664 (1998).
11. Grenfell, J. L. *et al.* *J. Geophys. Res.* **104**, 13771–13780 (1999).
12. O'Dowd, C. D. *et al.* *Geophys. Res. Lett.* **26**, 1707–1710 (1999).
13. Hoffmann, T. *et al.* *Geophys. Res. Lett.* **28**, 1949–1952 (2001).
14. Moyers, J. L. & Duce, R. A. *J. Geophys. Res.* **77**, 5229–5238 (1972).
15. Martens, C. J. *Geophys. Res.* **78**, 8867–8871 (1973).

Developmental genetics

Buffer zone

Massimo Pigliucci

Heat-shock proteins help to protect organisms from external stresses. The idea that they can also buffer against internal — genetic — variations has received support from studies of fruitflies and, now, of plants.

Living organisms are caught between a hammer and an anvil, evolutionarily speaking. On the one hand, they need to buffer the influences of genetic mutations and environmental stresses if they are to develop normally and maintain a coherent and functional form. On the other, stabilizing one's development too much may mean not being able to respond at all to changes in the environment and starting down the primrose path to extinction. On page 618 of this issue, Queitsch *et al.*¹ propose that, in plants, the balance between stability and the potential for change is made possible in part through a protein involved in 'heat-shock responses' in a wide variety of species, from plants to insects.

Heat-shock responses are a fundamental and widespread type of cellular defence against environmental stress. They have been studied for their effects on the fitness of organisms²; for their co-evolution with other cellular functions³; for their role in response to stresses not related to temperature⁴; and for the level of natural variation in the genes that encode the heat-shock proteins (Hsps)⁵, which mediate heat-shock responses. Notwithstanding the continuing debate⁶ about the actual function of these proteins, it is now clear that they are a complex family of gene products that are involved in protecting other proteins. Some Hsps are expressed continuously in the organism, whereas others are triggered by several environmentally harsh conditions (not only increases in temperature).

Given the ubiquity of these proteins and their role in protecting organisms from environmental changes, it makes sense to ask a more subtle question: can they also help to protect against disruptive genetic variations? After all, the systems that allow organisms to develop from a fertilized egg to the adult form have been honed over millions of years of evolution, and it is likely that a mutation in any of the tens of thousands of genes involved would disrupt the entire process, just as a severe environmental stress does. This idea

is rooted in the 1940s, in Waddington's classic studies^{7,8} of 'canalization' — the resistance of developing organisms to change when perturbed genetically or environmentally. More recently it has been suggested that, from the point of view of development, internal disturbances are simply another form of environmental change that needs to be properly 'canalized' to maintain a viable form (phenotype) tailored to specific functions⁹.

So can the Hsp proteins buffer genetic as well as environmental change? Rutherford and Lindquist¹⁰ first tested the idea of a connection between Hsp activity and genetic variation by looking at a popular animal model of developmental genetics — the fruitfly *Drosophila melanogaster*. The results were stunning. When the authors disrupted Hsp90, by either mutating or inhibiting it, phenotypic variation in nearly every structure of adult *D. melanogaster* ensued, with the details depending on the genetic background of the insects used (that is, on which other specific genes were present in each individual). This led the authors to conclude that *D. melanogaster* accumulates hidden genetic variation, which is somehow kept by Hsp90 from affecting the phenotype. If the function of Hsp90 is partly compromised, the buffer breaks and we can see previously 'unavailable' phenotypic variants.

Queitsch, Sangster and Lindquist¹ have now expanded this research to another model of developmental genetics, the plant *Arabidopsis thaliana*^{11,12}. These two species, *D. melanogaster* and *A. thaliana*, are of course very different in many ways. They have evolved separately over hundreds of millions of years. As one is a plant and the other an animal, they develop radically differently. And their breeding systems are not at all alike: fruit flies are obligatory 'out-crossers', meaning that they need a partner to produce offspring, whereas *A. thaliana* is mostly a 'selfer' — it fertilizes its own female gametes. Nonetheless, in *A. thaliana*, as in *D. melanogaster*, changes in Hsp90 release

previously hidden genetic variation, resulting in the production of novel phenotypes. These include altered leaf shapes, the accumulation of a purple pigment in hypocotyls (embryonic stems), and variations in hypocotyl length.

These latest findings¹ are also interesting for their differences from the previous results and for the potential follow-up that they make possible. For example, the range of phenotypes released by compromising Hsp90 function in *A. thaliana* seems less wide than the corresponding variation seen in *D. melanogaster* (in as much as it is possible to compare the phenotypes of insects and plants). This might be a result of the differences in developmental plasticity inherent to the two types of organism. Furthermore, by using sophisticated approaches to genetic analysis such as microarrays, it should be possible to further characterize the molecular basis of the phenotypic variation induced by blocking Hsp90 in *A. thaliana*.

More generally, there is no question about the importance of this sort of study for our understanding of the complex relationships between genes, environments and phenotypes. The mere ability to affect Hsp90 and possibly other similar proteins experimentally, thus producing an array of new phenotypes in a potentially wide range of organisms, provides evolutionary biologists with a powerful new research tool. It is less clear and more controversial whether these findings have long-term evolutionary implications, and the nature of the mechanisms that make the relationship between heat-shock proteins and hidden genetic variation possible is still unknown.

An obvious question is whether Hsp90 was actually selected to act as a 'capacitor' of morphological evolution, as it has been characterized by Rutherford and Lindquist¹⁰, or whether the storage of hidden genetic variation is instead a by-product of its normal physiological function. Although I would bet on the latter, only a phylogenetically informed comparative study, tracing the evolution of both the Hsp family itself and its phenotypic effects, will be able to shed light on this question.

Queitsch *et al.*¹ also suggest that, because of the breadth and strength of Hsp90's effects, it might have a large impact on evolutionary processes, and I am more inclined to agree on this point. But statements¹⁰ to the effect that the conditional release of hidden genetic variation may have allowed the rapid diversification of forms (radiations) that are occasionally seen in the fossil record might be too premature a connection between developmental genetics and palaeontology. Then again, it is exactly this sort of connection that will be needed to enlarge modern evolutionary theory to include molecular and developmental genetics⁹. ■

Massimo Pigliucci is in the Departments of Botany, and of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1100, USA.

e-mail: pigliucci@utk.edu

1. Queitsch, C., Sangster, T. A. & Lindquist, S. *Nature* **417**, 618–624 (2002); advance online publication, 12 May 2002 (DOI 10.1038/nature749).
2. Silberman, R. & Tatar, M. *Evolution* **55**, 2038–2045 (2001).
3. Knight, C. A. & Ackerly, D. D. *Am. J. Bot.* **88**, 411–418 (2001).
4. Kohler, H.-R., Zanger, M., Eckwert, H. & Einfeldt, I. *J. Evol. Biol.* **13**, 569–582 (2000).

5. Krebs, R. A. & Feder, M. E. *Evolution* **51**, 173–179 (1997).
6. Harrington, H. M., Dash, S., Dharmasiri, N. & Dharmasiri, S. *Aust. J. Plant Physiol.* **21**, 843–855 (1994).
7. Waddington, C. H. *Nature* **150**, 563–565 (1942).
8. Waddington, C. H. *Adv. Genet.* **10**, 257–290 (1961).
9. Schlichting, C. D. & Pigliucci, M. *Phenotypic Evolution, a Reaction Norm Perspective* (Sinauer, Sunderland, MA, 1998).
10. Rutherford, S. L. & Lindquist, S. *Nature* **396**, 336–342 (1998).
11. Anderson, M. & Roberts, J. A. (eds) *Arabidopsis* (Sheffield Academic Press, Sheffield, 1998).
12. Pigliucci, M. *Trends Plant Sci.* **3**, 485–489 (1998).

Condensed-matter physics

Indirect evidence

Andrew Millis

The ease with which electrons move through some metals depends on which direction they take. Watching electrons move along metal planes gives new insight, curiously, into how they move between them.

One of the fundamental issues in condensed-matter physics is conductivity, in other words how charge is transported through a material in response to an applied electric field. The essence of the conduction phenomenon in most materials is understood, thanks to work done from the 1930s to the 1950s by Bloch, Peierls and others. Over the past decade, however, the accepted wisdom has been challenged by measurements made on a wide range of 'strongly correlated metals', such as high-temperature copper-oxide superconductors, in which interactions between the conduction electrons dominate the physics.

On page 627 of this issue, T. Valla and colleagues (*Nature* **417**, 627–630; 2002) use a pleasingly indirect method to obtain a new insight into the conductivity of strongly correlated metals. Their work concerns electrically anisotropic materials (high-temperature superconductors are again an example) in which current can flow more easily in some directions than in others. Such materials pose a particular challenge: the temperature dependence of the conductivity varies markedly depending on the direction of current flow. Above a certain temperature — typically in the range 50–200 K, depending on the material — one finds 'metallic' behaviour (that is, conductivity decreases as the temperature rises) for the highly conducting directions and 'insulating' behaviour (conductivity increases as the temperature rises) for the poorly conducting directions. At lower temperatures the temperature dependence is the same for both directions of current flow, although the magnitudes of the conductivities differ greatly.

The existence of different temperature dependences for different directions of current flow is difficult to understand within the conventional Bloch–Peierls picture, in which conductivity is expressed as the product of electron velocity and the 'transport

time' (roughly, the mean time between electron-scattering events, which significantly degrade the current). Electron velocity is generically anisotropic but not temperature dependent, whereas the scattering time is generically temperature dependent but is not expected to depend strongly on the direction of current flow. The difficulty has stimulated much creative theoretical work but no generally accepted solution has yet emerged, in part because conductivity results were the only data that theorists had to work with.

Valla *et al.* have performed a photoemission experiment that sheds new light on this phenomenon: they studied two different layered materials whose crystal structure forms planes along which current can easily flow but in which the conductivity between planes is much smaller. The photoemission measurements involve shooting a beam of high-energy photons at a solid, then measuring the energy and in-plane momentum of emitted electrons. It seems remarkable that an experiment in which electrons are violently ejected from a solid can tell us something about the motion of electrons inside that solid. It is even more remarkable that Valla *et al.* have extracted important information about the ability of electrons to move between planes by studying the dependence of the photoelectron spectrum on in-plane electron momentum: Polonius's phrase from *Hamlet* — "by indirection find directions out" — comes to mind.

To understand what Valla *et al.* have shown, it is helpful to recall the basic theory of conductivity in metals. The quantum-mechanical effect of wavefunction hybridization enables electrons to move from one atom in a solid to another. In most circumstances, the timescale associated with this hybridization is very short, so that one can think of electrons as moving ballistically through the crystal. Metallic conductivity is

then associated with the presence of reasonably well-defined electron quasi-particle excitations. These quasi-particles carry unit charge and move ballistically through the crystal until they are scattered into another state, and can for all practical purposes be thought of just as normal electrons.

The mean time during which a quasi-particle propagates ballistically defines the quasi-particle lifetime. In the absence of imperfections in the crystal lattice, this lifetime is set by scattering off thermally excited electron-density fluctuations or off phonons (fluctuations of the lattice), and usually it diverges as the temperature decreases towards absolute zero. The quasi-particle lifetime is in fact shorter than the transport time (although Valla *et al.* conflate the two) for two reasons. First, a quasi-particle might be scattered out of one state and into another that is moving in substantially the same direction. Such forward-scattering processes do not degrade the current much. Second, a scattering process might cause a quasi-particle to decay into a combination of states that is not a quasi-particle but carries the same, or almost the same, current. The one-dimensional electron gas provides an extreme example of this: it possesses no stable excitation with quantum numbers that match those of the electron (so the quasi-particle lifetime is very short), but the possible stable excitations include states known as charged bosons that can move over very large distances without scattering, leading to a transport time that diverges strongly as the temperature approaches zero.

In quasi-two-dimensional materials, the presence of reasonably well-defined quasi-particle states can be inferred from the appearance of dispersing peaks in photoemission spectra (see for example in Fig. 1 on page 628). By comparing photoemission spectra with the interplane conductivity, Valla *et al.* demonstrate a striking correlation: quasi-particle peaks were present only in the low-temperature regime in which the interplane conductivity has a metallic temperature dependence. As the temperature is raised, the quasi-particle peak disappears at approximately the temperature at which the interplane conductivity crossed over from metallic to insulating behaviour.

At one level this is not surprising. The disappearance of the quasi-particle peak means that the quasi-particle lifetime has become very short. So a carrier is scattered many times in one plane before it can make a transition to the next — this is known as the 'incoherent transport regime'. It seems plausible that this strong scattering could lead to a non-metallic interplane conductivity, although a quantitative comparison of the data with the various theories of this effect has not yet been undertaken.

But if, thanks to the work of Valla *et al.*,

the interplane conductivity now seems less mysterious, the in-plane conductivity has become more so. The presence or absence of the quasi-particle peak is seemingly irrelevant to the in-plane conductivity, whose metallic temperature dependence extends smoothly through the crossover regime at which the quasi-particle peak vanishes. It may be that the in-plane conductivity is dominated by quasi-particles not observed in the photoemission experiment. But it is

more likely either that the scattering that destroys the quasi-particle is strongly peaked in the forward direction or, intriguingly, that the in-plane current is not carried by quasi-particles at all. Perhaps, to paraphrase Polonius, this measurement of incoherence has found a new type of coherence out. ■

Andrew Millis is in the Department of Physics, Columbia University, 538 West 120th Street, New York, New York 10027, USA.
e-mail: millis@phys.columbia.edu

Neurobiology

Glutamate receptors on the move

Morgan Sheng and Terunaga Nakagawa

Receptors for the neurotransmitter glutamate are more mobile than previously suspected. They meander about on the neuronal surface and become reversibly trapped at the junctions between neurons.

Neurons communicate with each other at specialized junctions known as synapses. In excitatory synapses of the brain, a signal-transmitting (presynaptic) neuron secretes glutamate — a chemical messenger that diffuses across the synaptic 'cleft' to bind to glutamate receptors concentrated in the postsynaptic membrane of the receiving neuron. A major subtype of glutamate receptor, the AMPA receptor, contains an ion channel that opens when glutamate binds to it, causing excitation of the post-synaptic neuron. The abundance of AMPA receptors in the postsynaptic membrane controls the magnitude of postsynaptic excitation and is a major factor in determining the strength of synaptic transmission. So it is clearly important to understand how these receptors accumulate in synapses. On page 649 of this issue,

Borgdorff and Choquet¹ reveal an unexpected dynamic behaviour of AMPA receptors (Fig. 1). Remarkably, the receptors wander around rapidly on the surface of neurons, stopping temporarily when they approach synapses. More interestingly, this 'lateral' mobility of AMPA receptors is controlled by signals that are known to regulate synaptic strength.

Modulation of the number of postsynaptic AMPA receptors is a potent mechanism used by neurons to change synaptic strength²⁻⁴. The resulting levels of AMPA receptors can be stable over long periods, so such a mechanism also offers a way to store information and memories in the brain. For these reasons, there is great interest in the dynamic distribution of AMPA receptors in neurons, particularly in how the receptors are delivered to synapses (leading to synaptic

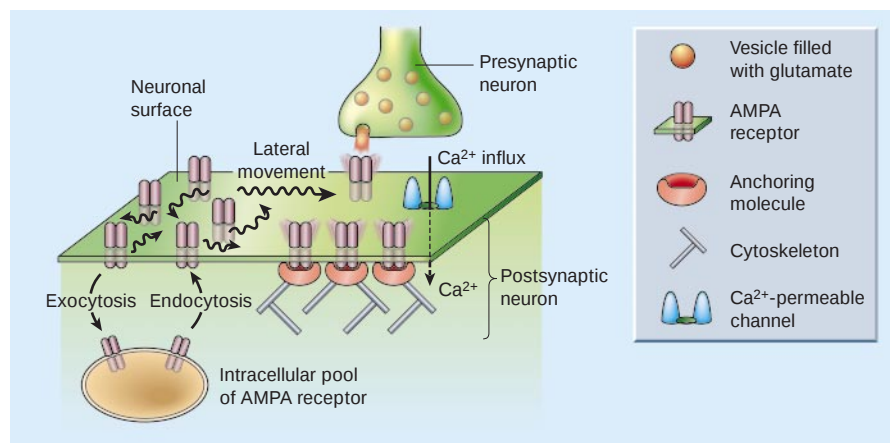


Figure 1 Wandering receptors. AMPA receptors, a subtype of receptor for the neurotransmitter glutamate, are found in postsynaptic neurons; they are ion channels that open in response to glutamate, which is released from presynaptic neurons. AMPA receptors can move between an intracellular pool and the neuronal surface, by the processes of exocytosis and endocytosis. As Borgdorff and Choquet¹ show, the receptors can also move by lateral diffusion in the plasma membrane until they are captured and immobilized by putative anchoring proteins concentrated in postsynaptic regions. AMPA receptors are also immobilized by increases in intracellular calcium levels, perhaps after calcium influx through postsynaptic calcium channels.



100 YEARS AGO

One of the latest departures of the experimental psychologist consists in prodding people with a pointed instrument when they are asleep to find out how much excitation is required before they begin to move, and how much it takes to wake them up. This method is embodied in a paper on "Experimental Investigations on the Depth of Sleep" by Drs. Sante de Sanctis and U. Neyroz, of Rome, a translation of which is given in the *Psychological Review* for May. The instrument employed is called a Griessbach ethesimeter (made by Brändli, of Basle), and may be used with either a sharp or blunt point. It measures the stimulus necessary to induce subconscious reaction, and that applied at the waking point... The [resulting] curves are all of zigzag form, and the experiments may perhaps suggest a practical application in the case of subjects who find it hard to wake in the morning, and who may overcome the difficulty by timing their sleep so that the waking point is at a minimum when they wish to rise.

From *Nature* 5 June 1902.

50 YEARS AGO

Progress in the field of sound recording was amply demonstrated in the annual exhibition of the British Sound Recording Association, held at the Waldorf Hotel, London, during May 17–18.... At long last the traditional wax cylinder for office dictating is being replaced by a plastic belt, which can be folded and sent through the post, as can also a paper disk carrying magnetic material, normally used when clamped on a turntable. In the field of commercial disk records, samples of the narrow-groove slow-speed records, which can play for a long time, have been in hand for some years, apart from talking-books for the blind; before the end of the year, two of the leading companies in this field will be making and distributing such disks. The 'battle of the revs', already in full swing in the United States, will soon be in full force in Britain also. Fortunately, many firms can supply reproducers with all three speeds, 78, 45 and 33 1/3 r.p.m., and all that the user has to take care of is the use of a needle with the correct radius of needle-tip.

From *Nature* 7 June 1952.

potentiation) and how they are removed (producing synaptic depression).

How can one study the dynamics of native AMPA receptors in neurons? The receptors alone are too small to see, even with a microscope, so Borgdorff and Choquet¹ coated tiny latex beads with antibodies that bind to the receptors. When applied to living neurons in culture, the beads adhere to AMPA receptors on the neuronal surface and act as microscopic signposts. In this way, lateral movements of the receptors can be tracked by video microscopy.

Using this approach, Borgdorff and Choquet found that AMPA receptors alternate abruptly between periods of rapid meandering (probably diffusion) on the neuronal outer membrane and periods of restricted motion within a submicrometre area ('confinement'). Confinement occurred mostly when the receptors were near synapses (which were detected with a specific dye). These findings imply that a zone of restricted diffusion exists in the neighbourhood of synapses. This 'perisynaptic' region probably contains specific anchoring proteins that bind to and tether AMPA receptors. The beads are too big to penetrate the narrow synaptic cleft, so a drawback of the approach is that it cannot sample true synaptic receptors. So we are left to presume that AMPA receptors located actually within the synapse are, like the perisynaptic receptors, also immobile.

Borgdorff and Choquet also observed that AMPA receptors sometimes moved from the vicinity of one dye-stained synaptic region to another. Whether this occurs *in vivo* is unclear. Normally, surface AMPA receptors undergo internalization (endocytosis), in which a patch of receptor-containing membrane is pinched off into the neuron as a tiny sac. Internalized receptors are then recycled back to the surface by secretion (exocytosis), in which the internal sacs fuse with the plasma membrane^{5,6}. The attachment of a latex bead to surface AMPA receptors is expected to prevent endocytosis; indeed, the authors found that abortive endocytosis seemed to account for a small minority of confinements. So perhaps the propensity to move from one synaptic region to another is exaggerated by the prolonged lifetime of bead-attached AMPA receptors on the neuronal surface.

More significantly, Borgdorff and Choquet showed that the surface mobility of AMPA receptors is influenced by intracellular calcium ions. The authors evoked a localized increase in calcium concentration in neurons by using a laser beam. In response, nearby AMPA receptors rapidly became immobilized, and this immobility persisted after the calcium spike. As postsynaptic calcium increases are a cardinal feature of excitatory synaptic transmission, calcium-dependent immobilization might explain why AMPA receptors accumulate at

synapses. A large postsynaptic increase in calcium concentration is also a well-known signal for strengthening of synapses, so, more interestingly, this could also be a mechanism for gathering more AMPA receptors in synapses to potentiate the postsynaptic response. Indeed, Borgdorff and Choquet observed that repeated calcium increases led to a local accumulation of AMPA receptors on the neuronal surface. But it remains to be seen whether surface AMPA receptors accumulate in response to postsynaptic calcium increases under more 'natural' conditions — that is, after synaptic stimulation.

Until now this field has focused on the idea that the delivery of AMPA receptors to synapses is regulated by exocytosis from internal compartments in the neuron^{6,7}. The study by Borgdorff and Choquet¹ points to another possibility, involving the lateral movement of surface receptors from extrasynaptic sites. The idea is supported by an earlier study of a mouse mutant called Stargazer, which shows neurological defects⁸. This study suggested that a critical step in synaptic delivery occurs after AMPA receptors reach the cell surface. More recently, another major subtype of glutamate receptor, the NMDA receptor, has been shown to move laterally from extrasynaptic to synaptic sites⁹. So a dynamic surface distribution might be a universal property of synaptic receptors.

What molecular mechanisms underlie the controlled lateral mobility of AMPA receptors? Presumably, it has something to do with calcium-regulated interactions between the receptors and intracellular anchoring proteins. More generally, how do synapses and synaptic circuits in the brain maintain consistent properties in the face of such molecular flux (assuming that it occurs *in vivo*)? The answer to this question is a long way off, but it is clear that the more we probe the molecular workings of synapses, the more we can understand their dynamic organization and their readiness for rapid change. ■

Morgan Sheng and Terunaga Nakagawa are at the Picower Center for Learning and Memory, RIKEN-MIT Neuroscience Research Center, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mails: msheng@mit.edu
terunaga@mit.edu

1. Borgdorff, A. J. & Choquet, D. *Nature* **417**, 649–653 (2002).
2. Malinow, R., Mainen, Z. F. & Hayashi, Y. *Curr. Opin. Neurobiol.* **10**, 352–357 (2000).
3. Carroll, R. C., Beattie, E. C., von Zastrow, M. & Malenka, R. C. *Nature Rev. Neurosci.* **2**, 315–324 (2001).
4. Sheng, M. & Lee, S. H. *Cell* **105**, 825–828 (2001).
5. Ehlers, M. D. *Neuron* **28**, 511–525 (2000).
6. Passafium, M., Piech, V. & Sheng, M. *Nature Neurosci.* **4**, 917–926 (2001).
7. Shi, S., Hayashi, Y., Esteban, J. A. & Malinow, R. *Cell* **105**, 331–343 (2001).
8. Chen, L. et al. *Nature* **408**, 936–943 (2000).
9. Tovar, K. R. & Westbrook, G. L. *Neuron* **34**, 255–264 (2002).

Astronomy

The Hoyle story

John Maddox

The achievements of Fred Hoyle, one of the twentieth century's great innovators in astronomy, were celebrated at a recent meeting. In many respects, it emerged, Hoyle's thinking was ahead of his time.

In 1952, Fred Hoyle delivered the first of the BBC's annual Reith Lectures under the title "Frontiers in Astronomy". The five talks, delivered in consecutive weeks, electrified his audience and won Hoyle an instant reputation as a popularizer of science. Half a century later, and some eight months after his death last year at the age of 86, a symposium under the same title* was convened to celebrate Hoyle's many achievements.

The meeting was held in Cambridge, UK, at the Institute of Astronomy, which Hoyle founded as the Institute of Theoretical Astronomy in 1966, and at St John's College, which elected him as a fellow while he was still a graduate student. The divided venue was a symbol of the spirit of the occasion: Cambridge's ambivalence towards Hoyle was matched only by Hoyle's ambivalence towards his old university. Another was that at the symposium even Hoyle's critics bit their tongues, acknowledging that he was "before his time". And C. Wickramasinghe (Cardiff Univ.), best known as Hoyle's collaborator on some of his later and more controversial ideas, roundly declared that "It is a myth that everything he did in his early years was brilliant and that everything he did after a certain date was rubbish". Nor was Wickramasinghe challenged on his dubious assertion that it is "now generally accepted" that the organic constituents of the first terrestrial living things were made in interstellar molecular clouds and were transported here by meteorites or comets.

The most telling account of Hoyle's precocity came from P. Solomon (State Univ. New York, Stony Brook), who had unearthed a paper published in 1943, in the *Proceedings of the Cambridge Philosophical Society*, when Hoyle was still a graduate student. The notion that molecular clouds cool by emitting radiation because of the rotational states of molecules was one of the ideas that first appeared there — and was then forgotten for almost 30 years. Solomon pointed out that, in the same paper, Hoyle used the idea to predict that the fragmentation of a cloud of gas (an early galaxy, say) into stellar nebulae would be constrained by the increasing opacity of collapsing regions, putting an upper limit on the mass of individual stars and restricting the numbers of Sun-like stars in galaxies (eventually pinned down in a paper dated 1949 to between 10^9 and 3×10^{11}).

*Frontiers in Astronomy, Institute of Astronomy/St John's College, Cambridge, UK, 16 April 2002.

The symposium was opened by W. Sargent (Caltech), himself a distinguished observer and Hoyle's principal adjutant on the Northern Hemisphere Telescope Review Committee. This was the body, set up in the mid-1960s, that eventually established the pattern of observing time for British astronomers in the past quarter of a century. Like the master of ceremonies at an awards dinner (except that there was no food and drink at 10:00 in the morning), Sargent rattled through Hoyle's lifetime of accomplishments.

Among them is the recognition of accretion (onto red giants in the first instance) as a general process affecting the smallest objects in the Universe (such as planets) and the largest — galaxies, our own included. Sargent stressed the importance of Hoyle's 1953 prediction that there is a long-lived excited state of the ^{12}C nucleus, the effect of which is to delay the conversion of ^{12}C into ^{16}O (by combination with ^4He nuclei), thus providing enough carbon in the cosmic mix to support carbon-based life. The prediction was confirmed by measurements by W. H. Fowler at

Caltech and is now widely cited as an illustration of the Anthropic Principle, the notion that the existence of living things implies that the Universe must have satisfied the pre-conditions for the existence of life as it is found, a biogenic supply of carbon included.

But, Sargent continued, Hoyle went on to explain the difference between supernovae of type I (with degenerate Fermi–Dirac cores) and type II (which implode), although in both cases Hoyle erred by underestimating the importance of neutrino emission in energy export.

Then there is the paper (known as B²FH), written with Geoffrey and Margaret Burbidge and W. H. Fowler on the synthesis in stars of the elements heavier than boron, which remains the universal framework for discussions of nucleosynthesis. Sargent went on to remind his audience that Hoyle had been the one to work out that stars 200 times as massive as the Sun could well have been formed in the early Universe and that he (with Wickramasinghe) first put the topic of interstellar dust on the astronomical agenda.

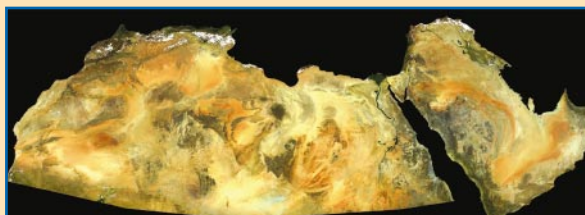
Sadly, of course, Hoyle was not present to receive the medal Sargent's talk seemed to presage. But others were. Margaret Burbidge (Univ. California, San Diego) gave a wistful account of how chance meetings at the Royal Astronomical Society in London, in Paris and at Cambridge grew into the collaboration that produced B²FH. And H. Bondi (Churchill College, Cambridge) spoke of the genesis of 'steady-state' theory, which, in

Earth science

African reflections

The outline of the geographical area shown here may be familiar, but the colouring won't be. This is north Africa and the Arabian peninsula, with the colours depicting the albedo of those parts of the region (almost all of it) that are arid and largely naked of vegetation. Albedo is a measure of the amount of solar radiation that is reflected back from a surface, into the atmosphere and space, and is therefore an essential parameter in climate modelling. Here, yellow corresponds to the highest albedos: the darker the shading, the lower the albedo (and the lower the amount of radiation reflected).

The albedo map comes from a paper by E. A. Tsvetinskaya *et al.* (*Geophys. Res. Lett.* **29**, 10.1029/2001GL014096; 2002), in which the authors



relate albedo measurements of unprecedented spectral detail taken by the Moderate-resolution Imaging Spectroradiometer (MODIS), on the Terra spacecraft, to digitized maps of regional soil and rock types, at a 1-km spatial resolution. On the albedo image, the brightest parts correspond to areas of sand dunes; the darkest parts correspond to soil types known as vertisols, cambisols, rendzinas and luvisols, which occur mostly in and around the Atlas Mountains, and to volcanic and intrusive rocks.

The reason that the authors chose north Africa and the Arabian peninsula for study is in part because this huge desert belt is one of the most reflective areas on Earth. The energy gradient between this and adjacent areas, which absorb radiation strongly, can produce distinctive regional circulations in the atmosphere. The detailed analysis of albedo by spatial variation in the surface type should mean that this kind of data can more easily be accommodated in climate modelling.

Tim Lincoln

contrast to the Big Bang, holds that expansion of the Universe can be accounted for by the continual creation of new matter. It was Bondi, along with Thomas Gold, who produced the first version of a steady-state cosmology; Hoyle joined them later and during the 1950s crossed swords publicly and acrimoniously with the late Martin Ryle. This great dispute, reflected Bondi, need not have happened if everybody had been more circumspect and respectful of the provisional

nature of newly gathered data. "Perhaps we were too ambitious", he acknowledged (in another context).

It fell to J. V. Narlikar (Inter-Univ. Centre for Astronomy and Astrophysics, Pune, India), another of Hoyle's former graduate students, to make the point that being ahead of your time is a dangerous business. The risk is that of being run over by the band-wagon that follows you. ■

John Maddox is Emeritus Editor of Nature.

Developmental neurobiology

Cortical liars

Seong-Seng Tan

Inhibitory cells known as interneurons constitute a significant proportion of the neurons in the neocortex of mammalian brains. As far as interneuron origins are concerned, humans may be the odd man out.

For more than a decade, the interneurons found in the neocortex of the human brain have been disturbing the peace of developmental neurobiologists. Interneurons represent roughly 20% of cortical neurons and are responsible for modulating the firing of the principal neurons known as projection neurons. They do this by producing the inhibitory neurotransmitter, GABA. But where exactly interneurons originate in the embryonic brain remains controversial. On page 645 of this issue¹, Rakic and colleagues provide evidence that one source of these cells for the neocortex that is not so evident in other mammals provides the majority of interneurons in humans. This is a striking conclusion, for it might help to explain the greater complexity of the human brain and what can go wrong in mental illness.

The controversy over the origins of cortical neurons hinges in part on the difference between radial and tangential movement of the cells from their sites of origin to their final home in the layers that make up the neocortex. Until about ten years ago, accepted wisdom was that cortical neurons migrate only radially: this is a comfortable notion, for it implies that the neocortex is built according to a carefully crafted blueprint with only local, outwards migration to create the cortical columns.

However, that notion was upset by the demonstration, in rodents, that a significant percentage of neurons migrate tangentially some distance from their point of origin². At that time, the identities of the tangentially migrating neurons were not known. But the fuss over migration patterns relates to the debate as to whether the embryonic neocortex is already carved into regionalized sectors³ or whether the emergence of functionally distinct areas in the adult is controlled by other mechanisms. Widespread and indeterminate tangential dispersion of

cells would, on the face of it, scramble any embryonic blueprint.

A bigger surprise emerged when second-generation cell-lineage studies revealed that tangential migration is associated with the development of interneurons, whereas projection neurons seemed to be the product of radial migration^{4,5}. At that time it was assumed that both kinds of neuron are born in the cortical ventricular zone (VZ), the main site of cell birth in the neocortex from which new cells cross an adjacent layer, the subventricular zone (SVZ), into the emerging cortical layers. An outline of embryonic brain geography is shown in Fig. 1.

But why would interneurons — which constitute a minority of cortical neurons, yet have the same developmental environment as projection neurons — migrate in a different direction? There were thought to be at least two possibilities. Either the migration pathway itself was somehow involved in turning cells into interneurons, or the progenitors residing in the VZ were already specified to make either interneurons (which for some reason migrated only tangentially) or projection neurons (which migrated only radially).

Unexpectedly, neither explanation was in itself correct. Instead, studies on mutants lacking the gene-transcription factors *Dlx1* and *Dlx2* showed that in these animals cortical interneurons were severely reduced in number. What is more, most interneurons expressing *Dlx1/2* were found to originate from a subcortical area⁶, the ganglionic eminence (shown in Fig. 1). This gives rise to subcortical structures such as the striatum, but the studies with mutant animals indicated that it also furnishes at least 75% of cortical interneurons⁷. Work with other species, including ferrets, produced similar observations, suggesting that this is an ancient mechanism of cortical development.

But what about interneurons in the embryonic human neocortex? The first hint that they might behave differently came from work on a different structure known as the thalamus, an important relay station for the neocortex. Letinic and Rakic⁸ discovered a migratory pathway that ferries interneurons generated in an external source, the ganglionic eminence, to the thalamus, and this pathway turned out to be absent from

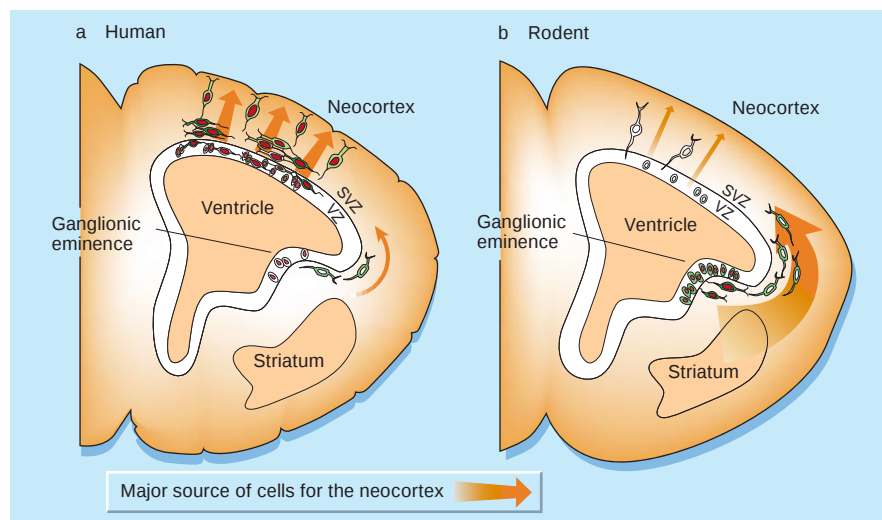


Figure 1 Interneuron origins in the embryonic brain. The graphic shows half of a cross-section through human (a) and rodent (b) forebrains. Rakic and colleagues¹ provide evidence that, during development, about 65% of interneurons in the human neocortex originate locally in the ventricular zone/subventricular zone (VZ/SVZ) of the neocortex. In rodents, by contrast, the main source of interneurons seems to be long-distance tangential migration of cells from the VZ of a subcortical area known as the ganglionic eminence. The other 35% of human interneurons seem to be produced by this route. In humans, interneurons originating from the cortical VZ/SVZ express both the transcription factors *Mash1* and *Dlx1/2* (cells with red nuclei); interneurons from the ganglionic eminence express only *Mash1* (white nuclei).

macaques and mice. Rakic and colleagues¹ now report that the ganglionic eminence also produces interneurons for the human neocortex, but the contribution is minor (35%). Instead, and most notably, they find that 65% of the interneurons arise locally from the cortical VZ/SVZ (Fig. 1a).

This source has two defining qualities. First, the cells migrate as chains, using each other as scaffolding to glide along; second, they express Mash1, a transcription-factor marker for interneuron progenitors. Surprisingly, Rakic and colleagues found that Mash1-positive progenitors reside mostly in the SVZ — a secondary germinal zone traditionally associated with the generation of another, non-neuronal cell type, glial cells. This implies that the cortical environment can nurture and support two types of developing neuron that are born at adjacent addresses (projection neurons in the VZ, and interneurons in the SVZ).

In the most telling experiments, Rakic and co-workers¹ infected cell progenitors in the cortical VZ/SVZ with a retrovirus carrying a gene tracer so that they could follow cell division and cell movement. By direct imaging of the infected cells, they showed that cells producing GABA also expressed the genes encoding Mash1 and *Dlx1/2*. But the idea that the interneuron source in the cortical VZ/SVZ is unique to humans needs to be counterbalanced by the finding that, in mice without *Dlx1/2*, a quarter of the cortical interneuron population still exists. The implication is that in other mammals interneurons are also generated locally in the cortical VZ.

Critics would also argue that the larger cortical interneuron source in humans might simply reflect a boosting of pre-existing developmental mechanisms. Indeed, Mash1 is expressed in certain cortical progenitors in mice⁹. Alternatively, the novel source of Mash1 progenitors in humans might arise from the evolutionary duplication of comparable cells in the ganglionic eminence. From work with rodents it seems that a single genetic switch is all that it takes — misexpression of Mash1 in cortical neurons results in their transformation to the interneuron type¹⁰. In this respect, laboratory studies with rodents might have recapitulated an event in human evolution.

Clarification of certain points raised by the new work¹ will be required. For instance, Rakic and colleagues show that interneurons actually migrate tangentially in the VZ/SVZ at first, followed by radial migration in the zone beyond that. In rodents, interneurons arising from ganglionic eminences behave in a similar manner, initially travelling close to the VZ as depicted in Fig. 1b, and subsequently changing tack to the radial direction¹¹. As Rakic and colleagues admit, they cannot exclude the possibility that some of the Mash1-positive cells in the VZ/SVZ might

have arrived from the ganglionic eminence at earlier embryonic stages and then continued to divide locally.

What about the evolutionary aspect? The primate neocortex is disproportionately large compared with that of other mammals, whereas the subcortical structures show only linear increases in size¹². Evolution of the relatively large neocortex would disturb the scaling of shared developmental processes between cortical and subcortical structures; and it would require a numerical increase in interneurons, which would also have longer distances to travel. Might the VZ of the ganglionic eminence (committed, of course, also to interneuron production for the thalamus) have been hard pressed to meet demand, resulting in the adaptational production of interneurons in the cortical VZ? There is also a temporal consideration. In humans, neuron migration into the neocortex occurs over a period of 200 days, compared with just 10 days in mice and 30 in ferrets. If interneurons were produced largely subcortically, as in rodents, it might be that the physical and molecular receptivity of the human cortical environment would become exhausted.

The debate about interneuron origins echoes lines from W. H. Auden's *Ode to the Diencephalon*: "How can you be quite so uncouth? After sharing the same skull for all these millennia, surely you should have discovered the cortical I is a compulsive liar." The cortical I — the interneuron cell type — has long been seemingly lying to observers about its behaviour. And then there is the medical connection. Malfunction of the GABA system is a hallmark of disorders such as schizophrenia¹³, to which defects in interneuron migration can be a predisposing factor. So could those defects be responsible for the 'cortical lies' perceived by the sufferer? Even when origins of interneurons have been sorted out, questions such as this will continue to disturb the peace of neurobiologists. ■

Seong-Seng Tan is in the Brain Development Laboratory, Howard Florey Institute, The University of Melbourne, Parkville 3010, Victoria, Australia.
e-mail: s.tan@hfi.unimelb.edu.au

1. Letinic, K., Zoncu, R. & Rakic, P. *Nature* **417**, 645–649 (2002).
2. Maricich, S. M., Gilmore, E. C. & Herrup, K. *Neuron* **31**, 175–178 (2001).
3. Rakic, P. *Proc. Natl Acad. Sci. USA* **92**, 11323–11327 (1995).
4. Mione, M. C., Cavanagh, J. F. R., Harris, B. & Parnavelas, J. G. *J. Neurosci.* **17**, 2018–2029 (1997).
5. Tan, S.-S. *et al. Neuron* **21**, 295–304 (1998).
6. Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. *Science* **278**, 474–476 (1997).
7. Marin, O. & Rubenstein, J. L. *Nature Rev. Neurosci.* **2**, 780–790 (2001).
8. Letinic, K. & Rakic, P. *Nature Neurosci.* **4**, 931–936 (2001).
9. Nieto, M., Schuurmans, C., Britz, O. & Guillemot, F. *Neuron* **29**, 401–413 (2001).
10. Fode, C. *et al. Genes Dev.* **14**, 67–80 (2000).
11. Nadarajah, B., Alifragis, P., Wong, R. O. & Parnavelas, J. G. *Nature Neurosci.* **5**, 218–224 (2002).
12. Finlay, B. L. & Darlington, R. B. *Science* **268**, 1578–1584 (1995).
13. Benes, F. M. & Berretta, S. *Neuropsychopharmacology* **25**, 1–27 (2001).

Daedalus

Toughened metal

Metallurgists are always trying to make stronger metals. One of their tricks is to mix molten metal with tiny solid particles, say of ceramic, to give a tough composite. The result is limited by poor adhesion between metal matrix and suspended particles; it also blunts cutting tools. Daedalus has a new plan.

Suppose, he says, you were to burn metal vapour, not in excess air or oxygen as a chemist would normally do it, but in excess metal vapour. The usual branched dendritic oxide particles would form; but on cooling they would get covered with metal from the surrounding excess vapour. Solid surfaces adhere strongly unless air or water vapour has coated them first (which is why solids hold together in the first place). So burning in excess metal vapour should give metal oxide particles pre-coated with metal.

DREADCO chemists are now trying it. They are burning metal vapours — aluminium, magnesium, zinc — in as little oxygen as possible; an inert-gas diluent may help. The cooled oxide particles should be superficially covered and wetted with metal.

This novel oxide should be easy to mix with melted metal. Unlike powdered ceramic, it should not float on top but, being already wetted with metal, will just seem to dissolve in the melt. The resulting 'Metox' composite will be greatly strengthened by its loading of fine oxide particles. Aluminium toughened by its own oxide should be more inextensible and much stronger than the metal itself. Magnesium and zinc should tell the same story. The chemists have many other metals to try — notably titanium. Sadly, metal oxide particles blunt the cutting tools used on metal. Metox composite will be best cast from the melt into its final form.

DREADCO's new Metox composites should be ideal for superchargers and turbine blades. Zinc binds conveniently firmly to iron (hence 'galvanized' zinc-plated steel sheets), so zinc-coated zinc oxide particles should mix readily with melted iron alloys. Daedalus also recalls how the graphite inclusions in cast iron make it remarkably silent. Silent Metox composites, or at least highly damped ones without the 'ring' of steel, would be a splendid contribution to our noisy civilization. But Daedalus's ultimate goal is a Metox with so much oxide that it is compatible with ceramics. It would be ideal for the pistons and cylinder-blocks of adiabatic engines, with no cooling system. They just run hot.

David Jones

Marine iguanas die from trace oil pollution

A near-miss ecological disaster still left a sinister aftermath for these giant lizards.

An oil tanker ran aground on the Galapagos island of San Cristóbal on 17 January 2001, spilling roughly three million litres of diesel and bunker oil. The slick started to spread westwards¹ and was dispersed by strong currents, so only a few marine animals were killed immediately as a result. Here we draw on long-term data sets gathered before the spill to show that a population of marine iguanas (*Amblyrhynchus cristatus*) on Santa Fe island suffered a massive 62% mortality in the year after the accident, due to a small amount of residual oil contamination in the sea. Another population on the more remote island of Genovesa was unaffected.

Environmental disasters such as oil fouling kill many organisms each year, but the subtle effects of low-level environmental contamination are rarely investigated². Although generally heralded as an ecological catastrophe narrowly averted³ at the Galapagos World Heritage site (Fig. 1), some oil still persisted around Santa Fe (Fig. 2a), where originally about 1 litre of

bunker oil came ashore per metre of beach; maximal oil concentrations reached 44 parts per million, which is low by international standards (L. Loughheed, G. Edgar and H.L.S., unpublished results).

The investigation of animal populations affected by environmental contamination generally begins in the aftermath of a spill so that the recovery process can be studied — as, for example, in the intertidal communities in Alaska after the *Exxon Valdez* accident⁴. However, in the years before the Galapagos oil spill, we had accrued long-term data sets on two island populations of marine iguanas⁵. We were therefore able to assess the effects of residual low-level oil contamination after the accident on one of these populations by comparing it with the other, whose habitat was unaffected.

We also compared circulating levels of the 'stress' hormone corticosterone in blood samples taken from marine iguanas on Santa Fe immediately before and shortly after the oil spill¹. We previously argued that



Figure 1 The tanker *Jessica* ran aground in the Galapagos archipelago in January 2001. Immediate damage to marine life was largely averted as the split oil was soon dispersed, but marine iguana populations in the vicinity fell by more than half during the following year.

HEIDI SNELL

exposed animals might suffer high mortality (Fig. 2b) as a result of the increased stress they experienced after the spill⁶.

Marine iguanas are herbivorous lizards that are endemic to the Galapagos and feed solely on algae in the intertidal or subtidal zone⁷. Their hindguts harbour specialized fermentation bacteria that break down algal cell walls to facilitate digestion⁸. Adult iguanas are essentially free of predators on both islands, so death is normally caused by senescence or by food shortages, which affect both populations to similar extents⁵. Mark-and-recapture data indicate that there were large population declines during the post-El Niño periods of 1983, 1988, 1992 and 1998 as a result of food shortages (Fig. 2c). These declines were followed by population recoveries during subsequent cold La Niña periods.

Because marine iguanas show strong site fidelity, their mortality can be measured by taking repeated censuses of study colonies^{5,6}. We collected this information from the two populations over periods of 20 years (Santa Fe, 90° 02' W, 0° 50' S) and 10 years (Genovesa island, 89° 59' W, 0° 19' N), respectively, before the oil spill, and for 11 months after the oil spill (until December 2001).

Despite the reportedly low degree of oil contamination, several lines of evidence indicate that the pronounced increase in mortality that we observed among the Santa Fe animals was related to the oil spill. First, although food conditions were good, we found an unexpectedly large number of skeletons along the shores of Santa Fe; no skeletons were found on Genovesa. Second, overall iguana numbers in the Santa Fe

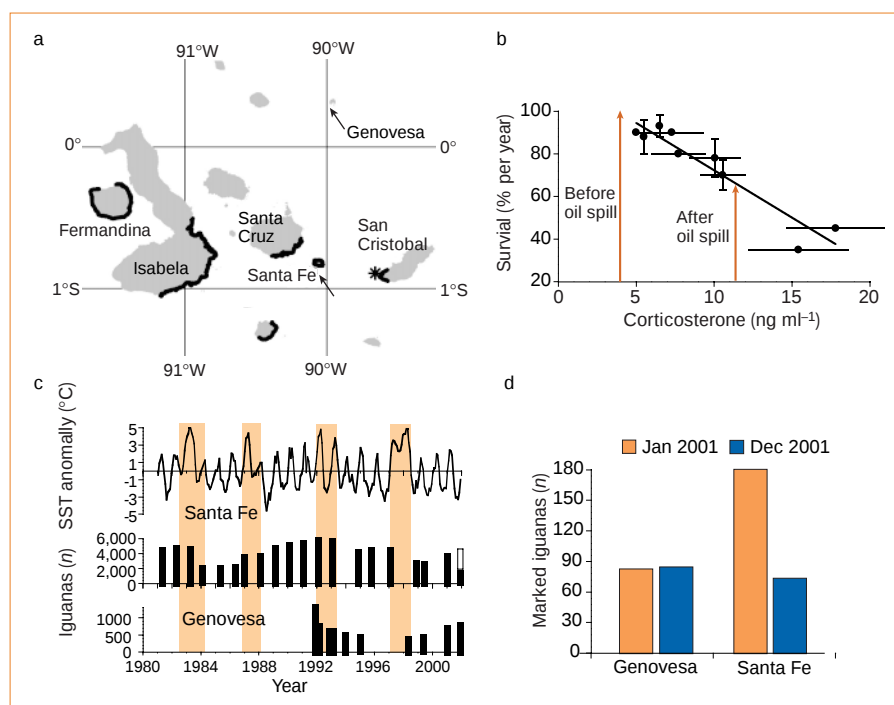


Figure 2 Effect on marine iguanas (*Amblyrhynchus cristatus*) of low-level oil contamination. **a**, Map of the Galapagos archipelago showing the site of the oil spill in January 2001 (asterisk) and the extent of oil contamination (thick lines) along the coastlines of the main islands until 15 February 2001 (data from coastal surveys; L. Loughheed, G. Edgar and H.L.S., unpublished results). Arrows show study sites. **b**, Corticosterone levels induced by 15 min of handling stress can be used to predict the survival of iguanas (redrawn from ref. 6). **c**, Anomalies in sea surface temperature (SST) in the Galapagos since 1980 (top), and population counts in study colonies of iguanas on the islands of Santa Fe and Genovesa. Iguana numbers generally decline during high-SST anomalies (El Niño periods). During 2001, there were more iguanas on Genovesa, as expected from hatchling recruitment during periods of low SST (high nutrient abundance), whereas Santa Fe showed a marked decline in numbers (white bar, predicted numbers; black bar, actual numbers in 2002). **d**, The number of permanently marked individuals recaptured during December 2001 was stable relative to January 2001 numbers on Genovesa, but decreased by 62% on Santa Fe.

study colony declined sharply (Fig. 2c). Third, we saw a significant reduction in the number of permanently marked iguanas on Santa Fe (Fig. 2d), whereas those on Genovesa all survived (62% local mortality of marked Santa Fe individuals; χ^2 -test, $P < 0.001$).

There are four possible explanations for this high mortality in the aftermath of the oil spill. The oil may have had a direct toxic effect either on the iguanas themselves⁹ or on the algae they consume⁴; the animals may have declined to eat because their food had been fouled; or their vital hindgut microsymbionts may have been poisoned so that the iguanas could no longer digest their food¹⁰.

As no deaths occurred immediately after the spill, and algal pastures and foraging were both normal in the 2 weeks after it, we infer that the fermenting endosymbionts in the iguanas' hindgut must be very sensitive to low-level environmental disturbance or contamination. This sensitivity probably compromises the digestive efficiency of affected iguanas⁷, causing their corticosterone concentrations to rise¹ and triggering an increase in mortality.

Our results illustrate the severe effects that low-level environmental contamination can have on wild animal populations¹¹. In this context, corticosterone levels are a reliable indicator of the induction of life-threatening stress^{6,11}, which correlates with the ensuing mass mortality on Santa Fe^{1,6}.

Ecology

Darwin's naturalization hypothesis challenged

Naturalized plants can have a significant ecological and economic impact¹, yet they comprise only a fraction of the plant species introduced into new areas by humans². Darwin proposed³ that introduced plant species will be less likely to establish a self-sustaining wild population in places with congeneric native species because the introduced plants have to compete with their close native relatives, or are more likely to be attacked by native herbivores or pathogens^{4,5}, a theory known as Darwin's naturalization hypothesis⁶. Here we analyse a complete list of seed-plant species that have been introduced to New Zealand and find that those with congeneric relatives are significantly more, not less, likely to naturalize — perhaps because they share with their native relatives traits that pre-adapt them to their new environment.

We obtained a list of all exotic angiosperm and gymnosperm species that have been introduced for cultivation into

Our findings warn against complacency over apparently low-impact contamination after environmental disasters in other wildlife areas, such as the Arctic National Wildlife Refuge in Alaska.

Martin Wikelski*, **Vanessa Wong***, **Brett Chevalier***, **Niels Rattenborg†**, **Howard L. Snell‡**

**Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA*

e-mail: wikelski@princeton.edu

†Department of Psychiatry, University of Wisconsin Medical School, Madison, Wisconsin 53719, USA

‡Charles Darwin Research Station, Galapagos, Ecuador, and Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA

- Wikelski, M., Romero, L. M. & Snell, H. L. *Science* **292**, 437–438 (2001).
- Paine, R. T. et al. *Annu. Rev. Ecol. Syst.* **27**, 197–235 (1996).
- Sea Technol. **42**, 74 (2001).
- van Tamelen, P. G., Stekoll, M. S. & Deysher, L. *Mar. Ecol. Prog. Ser.* **160**, 265–277 (1997).
- Wikelski, M. & Trillmich, F. *Evolution* **51**, 922–936 (1997).
- Romero, L. M. & Wikelski, M. *Proc. Natl Acad. Sci. USA* **98**, 7366–7370 (2001).
- Wikelski, M., Gall, B. & Trillmich, F. *Oecologia* **94**, 373–379 (1993).
- Mackie, R. I., Rycyk, M., Ruemmler, R. L., Aminov, R. I. & Wikelski, M. *Phys. Biochem. Zool.* (submitted).
- Prichard, A. K., Roby, D. D., Bowyer, R. T. & Duffy, L. K. *Chemosphere* **35**, 1531–1548 (1997).
- Ormseth, O. A. & Ben-David, A. *J. Comp. Physiol. B* **170**, 419–428 (2000).
- Fowler, G. S., Wingfield J. C. & Boersma, D. P. *Auk* **112**, 382–389 (1995).

Competing interests: financial, declared none; M.W. and the Galapagos National Park are in legal dispute with Petroecuador.

New Zealand, compiled using information obtained from commercial seed and plant catalogues, national agricultural research organizations, herbaria records, botanic gardens, the Ministry of Agriculture and Fisheries, private and commercial arboreta, specialist plant societies, and individuals.

Plant names were validated against the *Index Kewensis*⁷, synonyms were removed, and species were classified by family⁸. The resulting list comprised 24,774 species in 3,863 genera and 262 families. Of these, 1,769 species in 749 genera have become fully naturalized in New Zealand (these are defined as species that form populations that are self-maintained by seed or vegetative reproduction, or that occur repeatedly in wild or urban environments⁹).

Table 1 Classification of seed-plant genera introduced into New Zealand

		Genus contains native species	
		No	Yes
Genus contains	No	2,995	119
naturalized species	Yes	650	99
Naturalized genera		18%	45%

The 3,863 seed-plant genera introduced into New Zealand are cross-classified according to whether or not the genus contains at least one fully naturalized species, and whether it contains at least one native species.

We compiled a list of 1,511 native plant species that occur in genera with at least one introduced species. We separated the 3,863 introduced genera into those that contain at least one naturalized species and those that do not, and into those that contain at least one native species and those that do not (Table 1). Considerably fewer naturalized genera were found to contain native species (650 with no native species; 99 with at least one), but the naturalization rate was much higher among introduced genera that contain native species (45% in genera with native species; 18% in genera without).

To account for likely differences between families in naturalization rates, we fitted a generalized linear mixed model, using PROC GLMMIX in SAS¹⁰, with the naturalization rate per genus (that is, the number of naturalized species in a genus as a proportion of the total number of introduced species in a genus) as the response variable, a variable coding genera as containing at least one native species or not as a fixed-effect predictor, and a variable coding for family as a random effect.

This analysis revealed that, within families, genera containing at least one native species showed a significantly higher rate of naturalization than genera without native species ($t_{261} = 6.7$, $P < 0.0001$). Furthermore, for the 218 genera with at least one native species, the rate of naturalization was significantly higher in genera containing a greater number of native species ($t_{83} = 2.9$, $P < 0.005$, with family included as a random effect and number of native species per genus $\log_{10}$ -transformed).

Our discovery of a higher naturalization rate in introduced genera containing native species contradicts Darwin's naturalization hypothesis^{3–6}. One possible explanation, which was also considered by Darwin³, is that introduced species with native congeners are more likely to share features with them that allow survival in New Zealand, compared with introduced species that lack close relatives there. These shared traits may pre-adapt the plants to their new environment, helping to outweigh the potential disadvantage of stronger competition from close relatives^{11,12}.

Richard P. Duncan*, **Peter A. Williams†**

**Ecology and Entomology Group, Soil, Plant and Ecological Sciences Division, PO Box 84, Lincoln University, Canterbury, New Zealand*

e-mail: duncanr@lincoln.ac.nz

†Landcare Research, Private Bag 6, Nelson, New Zealand

- Vitousek, P. M., D'Antonio, C. M., Loope, L. L., Rejmánek, M. & Westbrooks, R. *NZ J. Ecol.* **21**, 1–16 (1997).
- Williamson, M. *Biological Invasions* (Chapman & Hall, London, 1996).
- Darwin, C. *On the Origin of Species* (Murray, London, 1859).
- Rejmánek, M. in *Invasive Species and Biodiversity Management* (eds Sandlund, O. T., Schei, P. J. & Viken, A.) 79–102 (Kluwer Academic, Dordrecht, 1999).

5. Mack, R. N. in *Weeds in a Changing World* (ed. Stirton, C. H.) 65–76 (British Crop Protection Council Symp. Proc. 64, Farnham, 1995).
6. Daehler, C. C. *Am. Nat.* **158**, 324–330 (2001).
7. Jackson, B. D. et al. *Index Kewensis Plantarum Phanerogamarum* (Clarendon, Oxford, 1895).
8. Mabberley, D. J. *The Plant Book: A Portable Dictionary of the Higher Plants* (Cambridge Univ. Press, Cambridge, 1987).
9. Heenan, P. B., Breitwieser, I., Glenn, D. S., de Lange, P. J. & Brownsey, P. J. *NZ J. Bot.* **36**, 155–162 (1998).
10. SAS Institute. *SAS/STAT Version 8* (SAS Institute, Cary, North Carolina, 1999).
11. Webb, C. O. *Am. Nat.* **156**, 145–155 (2000).
12. Williams, C. B. *J. Anim. Ecol.* **16**, 11–18 (1947).

Competing financial interests: declared none.

Palaeobiology

Calcification of early vertebrate cartilage

Hagfish and lampreys are unusual for modern vertebrates in that they have no jaws and their skeletons are neither calcified nor strengthened by collagen — the cartilaginous elements of their endoskeleton are composed of huge, clumped chondrocytes (cartilage cells). We have discovered that the cartilage in a 370-million-year-old jawless fish, *Euphanerops longaevus*, was extensively calcified, even

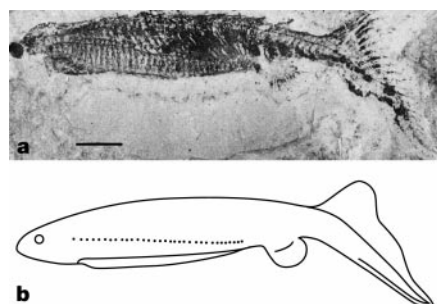


Figure 1 Young individual of the 370-million-year-old anaspid 'ostracoderm' *Euphanerops longaevus*. **a**, Specimen 2a from the Miguasha Museum of Natural History in Quebec, Canada. Scale bar, 10 mm. **b**, Reconstruction of the specimen in **a**. The specimen shows tarry imprints of cartilage that became calcified later in life.

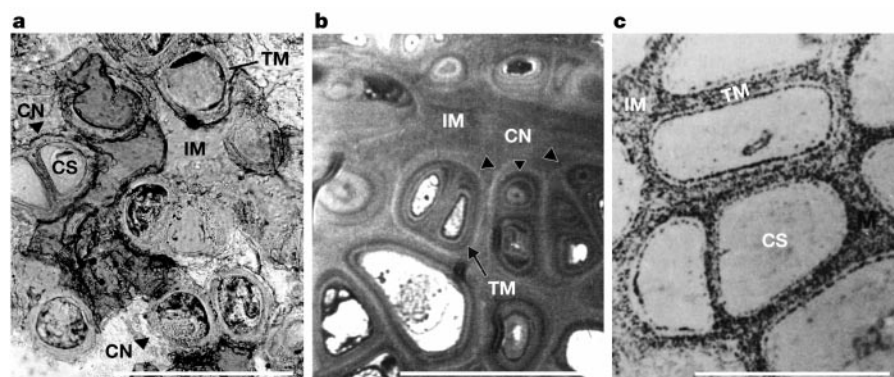


Figure 2 Micrographs of sections through an element of the vertebral column of a large specimen of *Euphanerops longaevus* (Miguasha Museum of Natural History no. 01-124) and through lamprey cartilage. **a**, *Euphanerops* displays large, rounded spaces for inclusion of chondrocytes, which are lined with calcified cartilage. **b**, Lamprey chondrocyte spaces in groups of two or four ('cell nests'), as in *Euphanerops*. **c**, Lamprey cartilage calcified *in vitro*, showing the same calcification pattern as in *Euphanerops*. Images in **b** and **c** are reproduced from ref. 4. CN, cell nest; CS, chondrocyte space; IM, intervening matrix; TM, territorial matrix. Scale bars, 50 μ m.

though its cellular organization was similar to the non-mineralized type found in lampreys. The calcification of this early lamprey-type cartilage differs from that seen in modern jawed vertebrates, and may represent a parallel evolutionary move towards a mineralized endoskeleton.

E. longaevus is an anaspid 'ostracoderm' (one of an ensemble of jawless fish aged 470–370 Myr) from the Late Devonian locality of Miguasha, Canada (Fig. 1). Although the endoskeleton of anaspids was previously thought not to have been calcified, that of large, recently discovered specimens of *Euphanerops* shows extensive calcification. All endoskeletal elements of these specimens show the same, foam-like aspect and are made up of minute, hollow, spherical bodies that are loosely attached either by their walls or by an intervening matrix (Fig. 2a). Thin sections reveal that these spherical bodies are sometimes divided into two chambers, and that their walls are composed of calcified cartilage with evidence of growth rings.

This structure is strikingly similar to that of lamprey cartilage, in which large chondrocyte cells, often in groups of two or four (cell 'nests'), are surrounded by a 'territorial' matrix with concentric rings that give rise to differential staining (Fig. 2b). In lamprey cartilage that is artificially calcified *in vitro*, calcium phosphate is deposited first in the territorial matrix (black dots in Fig. 2c) and then in the intervening matrix², generating the same pattern as is found in *Euphanerops*.

It is likely that the spherical bodies in *Euphanerops* contained chondrocytes, surrounded by calcified territorial matrix and held together loosely by a less-calcified intervening matrix. This process of calcification also occurs in the early stages of development of the higher jawed vertebrates (gnathostomes), but it must have occurred later in life in *Euphanerops*.

Ostracoderms are thought to be more closely related to gnathostomes than to

either lampreys or hagfish^{3,4}, but our discovery suggests that the cartilage of *Euphanerops* was structurally similar to that of lampreys, and that it may also have been non-collagenous. This does not necessarily imply a close relationship between anaspids and lampreys, as has been proposed⁵. There is no evidence that the calcification pattern that is found in *Euphanerops*, and which can be imposed *in vitro* on lamprey cartilage, is a precursor of the large-scale, globular calcification of cartilage and bone that is seen in more derived ostracoderms and in gnathostomes, although it may represent a parallel — but less successful — move towards developing a calcified endoskeleton.

Philippe Janvier*, Marius Arsenault†

*CNRS, UMR 8569, Muséum National d'Histoire Naturelle, 75231 Paris Cedex 05, France

e-mail: janvier@mnhn.fr

†Parc de Miguasha, CP 183, Nouvelle GOC2EO, Canada

1. Robson, P., Wright, G. M., Youson, J. H. & Keeley, F. W. *Comp. Biochem. Physiol.* **118B**, 71–78 (1997).
2. Langille, R. M. & Hall, B. K. *Acta Zool.* **74**, 31–41 (1993).
3. Janvier, P. *Palaeontology* **39**, 259–287 (1996).
4. Donoghue, P. C. J., Forey, P. L. & Aldridge, R. J. *Biol. Rev.* **75**, 191–251 (2000).
5. Janvier, P. *Early Vertebrates* (Clarendon, Oxford, 1996).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Microbial evolution

Antitoxin vaccines and pathogen virulence

In their mathematical description of vaccine-induced immunity¹, Gandon *et al.* predict that the host's immunity to a microbial toxin will lead to increased virulence of the pathogen without affecting its transmission. However, results obtained using vaccines against two additional bacterial pathogens are not consistent with this prediction. The model proposed by Gandon *et al.* may therefore be an oversimplification, with the outcome depending on the biological details of both the pathogen and the vaccine.

The equation given by Gandon *et al.*¹ implies that antitoxin immunity always selects for higher virulence by reducing the risk of host death and hence selecting for more virulent variants. The authors conclude that antitoxin vaccines are therefore worse from an evolutionary and epidemiological viewpoint than vaccines that induce other types of immunity. They argue that virulence evolves with the efficacy of antitoxin vaccines because this type of vaccine removes the cost of virulence (increased mortality) without affecting its benefit for the pathogen (increased transmission).

Results obtained using diphtheria and

pertussis vaccines that induce antitoxin immunity and that are widely used in human populations argue against these predictions. The introduction of diphtheria toxoid vaccine at the beginning of the twentieth century led to a huge reduction in the number of people carrying the virulent form of this pathogen and to the persistence of non-virulent forms of the bacterium²⁻⁴.

Diphtheria is caused by a toxin that is synthesized by *Corynebacterium diphtheriae*, which allows this bacterium to obtain nutrients when resources in the immediate vicinity are scarce. To produce the toxin, *C. diphtheriae* must carry a viral *tox* gene (*tox*⁺ strain). Toxin production therefore confers a competitive advantage — cases of frank diphtheria are more contagious than cases of asymptomatic infection.

However, toxin production also carries a metabolic cost. As the toxin is neutralized in people who are immunized with diphtheria toxoid, its production is a drain on the bacterium, which is therefore at a competitive disadvantage. Accordingly, diphtheria has vanished from areas with long-standing and thorough diphtheria-toxoid vaccination programmes, whereas the *tox*⁻ *C. diphtheriae* strain has persisted, a change that is attributable to the selective pressure exerted by the vaccine⁵.

A similar mechanism could explain the impact of the pertussis-vaccination programme implemented in Sweden with a vaccine containing only pertussis toxoid, which also induces antitoxin immunity. This vaccine was introduced in 1995 in 11 Swedish counties to vaccinate all children between 6 months and 14 years of age. Four years later, the result of this programme was a large reduction in hospitalized pertussis cases, not only in vaccinated but also in non-vaccinated children (that is, infants younger than 6 months old and children older than 14 years). This demonstrates once again that antitoxin immunity does affect pathogen transmission⁶⁻⁸.

Gandon *et al.* also argue that vaccines that counteract pathogen propagation may be less effective, as reduced transmission will elicit increased virulence. As we do not yet have an example of this type of vaccine for humans, we do not know whether this will be the case. This may be important for HIV vaccines⁹ as well as for malaria, but we suspect that the reduction in transmission of a pathogen that replicates on mucosal surfaces will outweigh any possible increases in endogenous virulence.

Benoit Soubeyrand*, Stanley A. Plotkin†

*Department of Medical Affairs, Aventis Pasteur, 69007 Lyon, France

†Department of Paediatrics, University of Pennsylvania, Pennsylvania 19104, and Aventis Pasteur, Doylestown, Pennsylvania 18901, USA
e-mail: stanley.plotkin@aventis.com

1. Gandon, S., Mackinnon, M. J., Nee, S. & Read, A. F. *Nature* **414**, 751–755 (2001).
2. Pappenheimer, A. M. in *Bacterial Vaccines* (ed. Germanier, R.) 1–36 (Academic, Orlando, 1984).
3. Galazka, A. J. *Infect. Dis.* **181** (suppl.), 2–9 (2001).
4. Bergamini, M. *et al. Epidemiol. Infect.* **125**, 105–112 (2000).
5. Ewald, P. W. in *Concepts in Vaccine Development* (ed. Kaufmann, S. H. E.) 1–25 (de Gruyter, Berlin, 1996).
6. Taranger, J. *et al. Clin. Infect. Dis.* **33**, 1004–1009 (2001).
7. Schneerson, R., Robbins, J. B., Taranger, J., Lagergard, T. & Trollfors, B. *Lancet* **348**, 1289–1292 (1996).
8. Mortimer, A. E. & Wharton, M. in *Vaccines* (eds Plotkin, S. A. & Orenstein, W. A.) 140–157 (Saunders, Philadelphia, 1999).
9. Amara, R. R. *et al. Science* **292**, 69–74 (2001).

Gandon *et al.* reply — Soubeyrand and Plotkin question our contention that antitoxin vaccines may select for greater pathogen virulence, arguing that this has not been borne out in real-life cases of diphtheria and pertussis, in which the widespread use of antitoxin vaccines has led to a reduced incidence of severe disease. They explain this success in terms of direct effects by the toxin on transmission that are both beneficial and costly. They argue that antitoxin vaccines have relieved the pathogen of the cost of high virulence due to host mortality (as we do too), but that these vaccines also maintain the metabolic cost of producing the toxin, helping natural selection to weed out the toxin producers.

In our model, we assume no such effects of toxin production — we envisage toxin production as an unavoidable, unhelpful side-effect of parasite replication, as seems to be the case in malaria. The apparent contradiction between our predictions and the observations cited by Soubeyrand and Plotkin is therefore due to differences in the life histories of different pathogens.

Our model can easily be extended to incorporate the costs and benefits of toxin production by modifying the pathogen's fitness function as follows:

$$R_0[\tau] = \frac{\beta[\alpha + (1-r)\tau]}{(\delta + \alpha + (1-r)\tau)} e^{-c\tau}$$

where τ is the level of toxin production, r is the efficacy of the antitoxin vaccine, $e^{-c\tau}$ is the cost function of toxin production, β represents parasite transmission as an increasing function of both toxin production and another component of disease-induced mortality, α , and δ is natural host

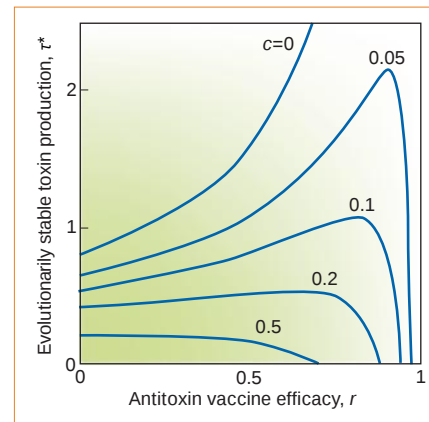


Figure 1 Evolutionarily stable toxin production, τ^* , plotted against antitoxin vaccine efficacy, r , for different toxin-production costs, c . Here it is assumed that all hosts are vaccinated, but similar results emerge for intermediate levels of vaccination coverage. The following transmission function was used: $\beta[\alpha + (1-r)\tau] = b_1(\alpha + (1-r)\tau)^{b_2}$. Parameter values: $b_1 = 1$, $b_2 = 0.5$, $\delta = 1$, $\alpha = 0.2$.

mortality. Maximizing fitness yields the evolutionarily stable toxin production, τ^* , shown in Fig. 1. When the cost of toxin production is zero (as is assumed in our original model), virulence increases with vaccine efficacy. When the cost of toxin production is high, however, it counteracts the toxin's benefit to transmission, in which case optimal toxin production decreases with vaccine efficacy.

Figure 1 also shows that whereas highly effective antitoxin vaccines select for lower toxin production, imperfect vaccines can select for higher toxin production, which supports our argument that the use of imperfect vaccines can have negative consequences. The examples provided by Soubeyrand and Plotkin emphasize the need to understand how virulence and transmission relate to pathogen fitness for each disease of interest. Virulence evolution can occur in response to vaccination and other increases in host defence, both in positive ways, as Soubeyrand and Plotkin argue has occurred for diphtheria and pertussis, and in negative ways, as others have argued may be the case in Marek's disease¹ and myxomatosis².

Sylvain Gandon, Margaret J. Mackinnon, Sean Nee, Andrew F. Read

*Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK
e-mail: sylvain.gandon@ed.ac.uk*

1. Witter, R. L. *Acta Vet. Hung.* **45**, 251–266 (1997).
2. Fenner, F. & Fantini, B. *Biological Control of Vertebrate Pests: The History of Myxomatosis — an Experiment in Evolution* (CABI, Wallingford, UK, 1999).

brieF communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com/nature) or on request from nature@nature.com

Computational and evolutionary aspects of language

Martin A. Nowak*, Natalia L. Komarova*† & Partha Niyogi‡

* Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA

† Department of Mathematics, University of Leeds, Leeds LS2 9JT, UK

‡ Department of Computer Science, University of Chicago, Chicago, Illinois 60637, USA

Language is our legacy. It is the main evolutionary contribution of humans, and perhaps the most interesting trait that has emerged in the past 500 million years. Understanding how darwinian evolution gives rise to human language requires the integration of formal language theory, learning theory and evolutionary dynamics. Formal language theory provides a mathematical description of language and grammar. Learning theory formalizes the task of language acquisition—it can be shown that no procedure can learn an unrestricted set of languages. Universal grammar specifies the restricted set of languages learnable by the human brain. Evolutionary dynamics can be formulated to describe the cultural evolution of language and the biological evolution of universal grammar.

Biology uses generative systems. Genomes consist of an alphabet of four nucleotides, which, together with certain rules for how to produce proteins and organize cells, generates an unlimited variety of living organisms. For more than 3 billion years, evolution of life on Earth was restricted to using this generative system. Only very recently another generative system emerged, which led to a new mode of evolution. This other system is human language. It enables us to transfer unlimited non-genetic information among individuals, and it gives rise to cultural evolution.

Currently there are many efforts to bring linguistic inquiry into contact with several areas of biology including evolution^{1–11}, genetics^{12–14}, neurobiology^{15,16} and animal behaviour^{17–20}. The aim of this Review is to formulate a synthesis of formal language theory^{21,22}, learning theory^{23–28} and evolutionary dynamics in a manner that is useful for people from various disciplines. We will address the following questions: What is language? What is grammar? What is learning? How does a child learn language? What is the difference between learning language and learning other generative systems? In what sense is there a logical necessity for genetically determined components of human language, such as ‘universal grammar’? Finally, we will discuss how formal language theory and learning theory can be extended to study language as a biological phenomenon, as a product of evolution.

Formal language theory

Language is a mode of communication, a crucial part of human behaviour and a cultural object defining our social identity. There is also a fundamental aspect of human language that makes it amenable to formal analysis: linguistic structures consist of smaller units that are grouped together according to certain rules.

The combinatorial sequencing of small units into bigger structures occurs at several different levels. Phonemes form syllables and words. Words form phrases and sentences. The rules for such groupings are not arbitrary. Any native English speaker recognizes that the sentence ‘He ran from there with his money’ obeys the rules of English, while ‘He his money with there from ran’ does not. In Bengali the reverse is true.

Individual languages have specific rules. Certain word orders are admissible in one language but not in another. In some languages, word order is relatively free but case marking is pronounced. There are always specific rules that generate valid or meaningful linguistic structures. Much of modern linguistic theory proceeds from this

insight. The area of mathematics and computer science called formal language theory provides a mathematical machinery for dealing with such phenomena.

What is language?

An alphabet is a set containing a finite number of symbols. Possible alphabets for natural languages are the set of all phonemes or the set of all words of a language. For these two choices one obtains formal languages on different levels, but the mathematical principles are the same. Without loss of generality, we can consider the binary alphabet, $\{0,1\}$, by enumerating the actual alphabet in binary code.

A sentence is defined as a string of symbols. The set of all sentences over the binary alphabet is $\{0,1,00,01,10,11,000,\dots\}$. There are infinitely many sentences, as many as integers; the set of all sentences is ‘countable’.

A language is a set of sentences. Among all possible sentences

An **alphabet** is a set of symbols:

$\{0,1\}$

Sentences are strings of symbols:

0,1,00,01,10,1,...

A **language** is a set of sentences:

$L = \{000,0100,0010,\dots\}$

A **grammar** is a finite list of rules defining a language.

$S \rightarrow 0A$	$B \rightarrow 1B$
$A \rightarrow 1A$	$B \rightarrow 0F$
$A \rightarrow 0B$	$F \rightarrow \epsilon$

Figure 1 The basic objects of formal language theory are alphabets, sentences, languages and grammars. Grammars consist of rewrite rules: a particular string can be rewritten as another string. Such rules contain symbols of the alphabet (here 0 and 1), and so-called ‘non-terminals’ (here S, A, B and F), and a null-element, ϵ . The grammar in this figure works as follows: each sentence begins with the symbol S. S is rewritten as 0A. Now there are two choices: A can be rewritten as 1A or 0B. B can be rewritten as 1B or 0F. F always goes to ϵ . This grammar generates sentences of the form 01^m01^n0 , which means that every sentence begins with 0, followed by a sequence of m 1s, followed by a 0, followed by a sequence of n 1s, followed by 0.

some are part of the language and some are not. A finite language contains a finite number of sentences. An infinite language contains an infinite number of sentences. There are infinitely many finite languages, as many as integers. There are infinitely many infinite languages, as many as real numbers; they are not countable. Hence, the set of all languages is not countable.

What is grammar?

A grammar is a finite list of rules specifying a language. A grammar is expressed in terms of 'rewrite rules': a certain string can be rewritten as another string. Strings contain elements of the alphabet together with 'non-terminals', which are place holders. After iterated application of the rewrite rules, the final string will only contain symbols of the alphabet. Figures 1 and 2 give examples of grammars.

There are countably infinitely many grammars; any finite list of

rewrite rules can be encoded by an integer. As there are uncountably many languages, only a small subset of them can be described by a grammar. These languages are called 'computable'.

Languages, grammars and machines

There is a correspondence between languages, grammars and machines. 'Regular' languages are generated by finite-state grammars, which are equivalent to finite-state automata. Finite-state automata have a start, a finite number of intermediate states and a finish. Whenever the machine jumps from one state to the next, it emits an element of the alphabet. A particular run from start to finish produces a sentence. There are many different runs from start to finish, hence there are many different sentences. If a finite-state machine contains at least one loop, then it can generate infinitely many sentences. Finite-state automata accept all finite languages

Box 1

Statistical learning theory and other approaches

Classical learning theory as formulated by Gold²³ makes a number of somewhat problematic assumptions: (1) the learner has to identify the target language exactly; (2) the learner receives only positive examples; (3) the learner has access to an arbitrarily large number of examples; and (4) the learner is not limited by any consideration of computational complexity. Assumptions (1) and (2) are restrictive: relaxing these assumptions will enable particular learners to succeed on larger sets of languages. Assumptions (3) and (4) are unrestrictive: relaxing these assumptions will reduce the set of learnable languages. In fact, each assumption has been removed in various approaches to learning, but the essential conclusion remains the same: no algorithm can learn an unrestricted set of languages.

Perhaps the most significant extension of the classical framework is statistical learning theory^{24–26}. Here, the learner is required to converge approximately to the right language with high probability. Let us consider the following objects that play an important role in the basic framework of statistical learning:

- Languages and indicator functions. Let $\mathcal{L}$ be a set of languages. Every language L of this set defines an indicator function $1_L(s)$ that takes the value 1 if a sentence s is in L and 0 otherwise.
- Linguistic examples. Linguistic examples are provided to the learner according to some distribution P on the set of all sentences. The learner receives both positive and negative examples.
- Distances between languages. The probability measure P has a dual role. In addition to providing sentences, it also defines the distance between two languages, $d(L_1, L_2) = \sum_s |1_{L_1}(s) - 1_{L_2}(s)|P(s)$.
- Learnability. The criterion for learnability is specified as follows. Assume some language $L \in \mathcal{L}$ is the target language. The learner receives a collection of positive and negative example sentences according to the distribution P . On the basis of these empirical data, the learner guesses a language $L_N \in \mathcal{L}$ after N examples have been received. The target L is said to be learnable if for all $\epsilon > 0$ the probability that the distance between L_N and L is greater than ϵ converges to 0 as $N \rightarrow \infty$. As a consequence, there exists a finite number of example sentences, such that the probability is greater than $1 - \delta$ (where δ is a small number) that the learner's current estimate is within an ϵ -approximation of the target language.

A deep result, originally due to Vapnik and Chervonenkis²⁴ and elaborated since, states that a set of languages is learnable if and only if it has finite VC dimension. The VC dimension is a combinatorial measure of the complexity of a set of languages. Thus if the set of possible languages is completely arbitrary (and therefore has infinite VC dimension), learning is not possible. It can be shown that the set

of all regular languages (even the set of all finite languages) has infinite VC dimension and hence cannot be learned by any procedure in the framework of statistical learning theory. Subsets of regular languages that are generated by finite-state automata with n states, however, have finite VC dimension, and one can estimate bounds on the number of sample sentences that are needed for learning.

Statistical learning theory in the VC framework removes assumption (1), (2) and (3) of the Gold framework: it does not ask for convergence to exactly the right language, the learner receives positive and negative examples, and the learning process has to end after a certain number of examples. The theory provides bounds for how many example sentences are needed to converge approximately to the right language with high probability; this is the concept of informational complexity.

Valiant²⁵ also added considerations of computational complexity, thereby removing assumption (4) of the Gold framework: the learner is required to approximate the target grammar with high confidence using an efficient algorithm. Consequently, there are sets of languages that are learnable in principle (have finite VC dimension), but no algorithm can do this in polynomial time. (Computer scientists consider a problem 'intractable' if no algorithm can solve the problem in polynomial time, which means in a number of time steps that is proportional to the size of the input raised to some power.)

Some other models of learning deserve mention. For example, in one form of query-based learning, the learner is allowed to ask whether a particular sentence is in the target language or not. In this model, regular languages can be learned in polynomial time⁷², but context-free languages cannot⁷³. Other query-based models of learning, with varying degrees of psychological plausibility, have been considered, and none permit all languages to be learnable⁷⁴.

Another model of learning follows the classical Gold framework but only requires the learner to identify the target language on almost all texts. Texts are assumed to be generated by a probabilistic process, where sentences are drawn according to some measure μ on the target language. If the learner is to identify the target grammar in a distribution free fashion (that is independent of μ), then the set of learnable languages is no larger than those that are learnable in the classical Gold framework. If constraints are put on the family of measures μ then the set of learnable languages can be enlarged. Such constraints act like a probabilistic form of universal grammar.

In summary, all extensions of learning theory underline the necessity of specific restrictions.

and some infinite languages.

'Context-free' languages are generated by context-free grammars, which can be implemented by push-down automata. These are computers with a single memory stack: at any one time they have only access to the top register of their memory.

'Context-sensitive' languages are generated by context-sensitive

Box 2

Language as mapping between sound and meaning

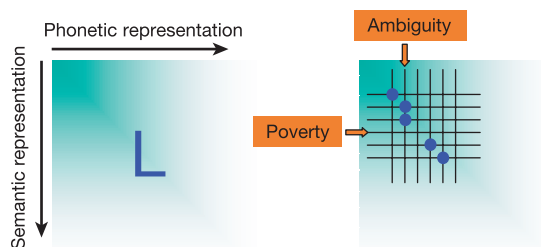
The mathematical formalism of language can be extended to include aspects of communicative behaviour and performance. A language can be seen as a (possibly infinite) matrix, L , that specifies mappings between phonetic forms and semantic forms that are 'sound' and 'meaning' insofar as they are linguistically determined (see Box 2 figure). This matrix defines linguistic competence. For evaluating communicative success, we also need to describe linguistic performance. We assume that the matrix, L , leads to a pair of matrices, P and Q , determining speaking and hearing. The element p_{ij} denotes the probability for a speaker of L_i to use sound j for encoding meaning i . The element q_{ji} denotes the probability for a hearer to decode sound j as meaning i . A language may not encode all meanings and may not make use of all possible sounds.

Next, we introduce a measure, σ , on the set of all meanings. Let us denote by σ_i the probability that communication is about meaning i . The measure, σ , depends among other things on the environment, behaviour and phenotype of the individuals. It also defines which meanings are more relevant than others.

The probability that a speaker of L_i generates a sound which is understood by a hearer using L_j is given by $a_{ij} = \sum_{\sigma} \sigma_i p_{ij} q_{j\sigma}$. The communicative pay-off between L_i and L_j can be defined as $F_{ij} = \frac{1}{2}(a_{ij} + a_{ji})$. The intrinsic communicative pay-off of L_i is $F_{ii} = a_{ii}$.

In this framework, communicative pay-off is a number between 0 and 1. A pay-off of less than 1 arises as a consequence of ambiguity and poverty. Ambiguity, α_i , of language L_i is the loss of communicative capacity that arises if individual sounds are linked to more than one meaning. Poverty, β_i , is the fraction of meanings (measured by σ) that are not part of L_i . The communicative capacity of L_i can be written as $F_{ii} = (1 - \alpha_i) \times (1 - \beta_i)$.

For language acquisition we need a measure for the similarity, s_{ij} , between languages L_i and L_j . A possibility is $s_{ij} = \sum_{\sigma} \sigma_i p_{ij} q_{j\sigma} / \sum_{\sigma} \sigma_i p_{ij}$ denoting the probability that a sound generated by a speaker of L_i is correctly interpreted by a hearer using L_j . In this context, the similarity between L_i and L_j declines because of ambiguity, but not because of poverty. Ambiguity implies that a learner holding the correct hypothesis might think he is wrong and change his hypothesis. This has consequences for the notion of consistent learning; a consistent learner does not change his hypothesis if he already holds the correct hypothesis.



grammars. For each of these languages there exists a Turing machine, which can decide for every sentence whether it is part of the language or not. A Turing machine embodies the theoretical concept of a digital computer with infinite memory.

Computable languages are described by 'phrase structure' grammars that have unrestricted rewrite rules. For each computable language, there exists a Turing machine that can identify every sentence that is part of the language. If, however, the Turing machine receives as input a sentence which does not belong to the language, then it might compute forever. Hence, the Turing machine cannot always decide whether a sentence is part of the language or not.

Figure 3 shows the Chomsky hierarchy: finite-state grammars are a subset of context-free grammars, which are a subset of context-sensitive grammars, which are a subset of phrase-structure grammars, which are Turing complete.

The structure of natural languages

With the introduction of the Chomsky hierarchy, there was some interest in placing natural languages within this scheme. Natural languages are infinite: it is not possible to imagine a finite list that contains all English sentences. Furthermore, finite-state grammars are inadequate for natural languages. Such grammars are unable to represent long-range dependencies of the form 'if... then'. The string of words between 'if' and 'then' could be arbitrarily long, and could itself contain more paired if-then constructions. Such pairings relate to rules that generate strings of the form $0^n 1^n$, which require context-free grammars (Fig. 2). There is a continuing debate whether context-free grammars are adequate for natural languages, or whether more complex grammars need to be evoked^{29,30}.

The fundamental structures of natural languages are trees. The nodes represent phrases that can be composed of other phrases in a recursive manner. A tree is a 'derivation' of a sentence within the rule system of a particular grammar. The interpretation of a sentence depends on the underlying tree structure. Ambiguity arises if more than one tree can be associated with a given sentence. One can also define grammars that directly specify which trees are acceptable for a given language. Much of modern syntactic theory

Finite state	Context free	Context sensitive
$S \rightarrow 0S$	$S \rightarrow 0S1$	$S \rightarrow 0AS2$
$S \rightarrow A$	$S \rightarrow \epsilon$	$S \rightarrow 012$
$A \rightarrow 1A$	$L = 0^n 1^n$	$A0 \rightarrow 0A$
$A \rightarrow \epsilon$		$A1 \rightarrow 11$
$L = 0^n 1^n$		$L = 0^n 1^n 2^n$

Figure 2 Three grammars and their corresponding languages. Finite-state grammars have rewrite rules of the form: a single non-terminal (on the left) is rewritten as a single terminal possibly followed by a non-terminal (on the right). The finite-state grammar, in this figure, generates the regular language $0^n 1^n$; a valid sentence is any sequence of 0s followed by any sequence of 1s. A context-free grammar admits rewrite rules of the form: a single non-terminal is rewritten as an arbitrary string of terminals and non-terminals. The context-free grammar in this figure generates the language $0^n 1^n$; a valid sentence is a sequence of 0s followed by the same number of 1s. There is no finite-state grammar that could generate this language. A context-sensitive grammar admits rewrite rules of the form $\alpha A \beta \rightarrow \alpha \gamma \beta$. Here α , β and γ are strings of terminals and non-terminals. Although α and β may be empty, γ must be non-empty. The important restriction on rewrite rules of context-sensitive grammars is that the complete string on the right must be at least as long as the complete string on the left. The context-sensitive grammar, in this figure, generates the language $0^n 1^n 2^n$. There is no context-free grammar that could generate this language.

deals with such grammars^{31–34}, which are, of course, also part of the Chomsky hierarchy, and the results of learning theory, to be discussed now, apply to them.

Learning theory

Learning is inductive inference. The learner is presented with data and has to infer the rules that generate these data. The difference between ‘learning’ and ‘memorization’ is the ability to generalize beyond one’s own experience to novel circumstances. In the context of language, the child will generalize to novel sentences never heard before. Any person can produce and understand sentences that are not part of his previous linguistic experience. Learning theory describes the mathematics of learning with the aim of outlining conditions for successful generalization.

The paradox of language acquisition

Children learn their native language by hearing grammatical sentences from their parents or others. From this ‘environmental input’, children construct an internal representation of the underlying grammar. Children are not told the grammatical rules. Neither children nor adults are ever aware of the grammatical rules that specify their own language.

Chomsky pointed out that the environmental input available to the child does not uniquely specify the grammatical rules³⁵. This phenomenon is known as ‘poverty of stimulus’³⁶. ‘The paradox of language acquisition’ is that children of the same speech community reliably grow up to speak the same language³⁷. The proposed solution is that children learn the correct grammar by choosing from a restricted set of candidate grammars. The ‘theory’ of this restricted set is ‘universal grammar’ (UG). Formally, UG is not a grammar, but a theory of a collection of grammars.

The concept of an innate, genetically determined UG was controversial when introduced some 40 years ago and has remained so. The mathematical approach of learning theory, however, can explain in what sense UG is a logical necessity.

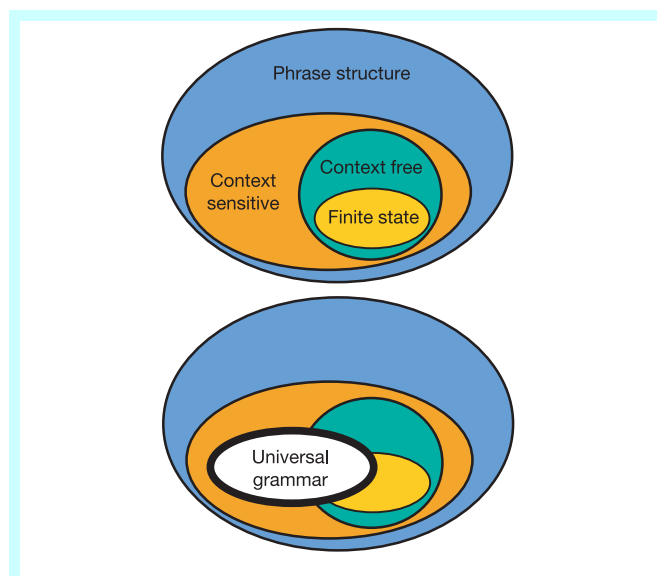


Figure 3 The Chomsky hierarchy and the logical necessity of universal grammar. Finite-state grammars are a subset of context-free grammars, which are a subset of context-sensitive grammars, which are a subset of phrase-structure grammars, which represent all possible grammars. Natural languages are considered to be more powerful than regular languages. The crucial result of learning theory is that there exists no procedure that could learn an unrestricted set of languages; in most approaches, even the class of regular languages is not learnable. The human brain has a procedure for learning language, but this procedure can only learn a restricted set of languages. Universal grammar is the theory of this restricted set.

Learnability

Imagine a speaker–hearer pair. The speaker uses grammar, G , to construct sentences of language L . The hearer receives sentences and should after some time be able to use grammar G to construct other sentences of L . Mathematically speaking, the hearer is described by an algorithm (or more generally, a function), A , which takes a list of sentences as input and generates a language as output.

Let us introduce the notion of a ‘text’ as a list of sentences. Specifically, text T of language L is an infinite list of sentences of L with each sentence of L occurring at least once. Text T_N contains the first N sentences of T . We say that language L is learnable by algorithm A if for each T of L there exists a number M such that for all $N > M$ we have $A(T_N) = L$. This means that, given enough sentences as input, the algorithm will provide the correct language as output.

Furthermore, a set of languages is learnable by an algorithm if each language of this set is learnable. We are interested in what set of languages, $\mathcal{L} = \{L_1, L_2, \dots\}$, can be learned by a given algorithm.

A key result of learning theory, Gold’s theorem²³, implies there exists no algorithm that can learn the set of regular languages. As a consequence, no algorithm can learn a set of languages that contains the set of regular languages, such as the set of context-free languages, context-sensitive languages or computable languages.

Gold’s theorem formally states there exists no algorithm that can learn a set of ‘super-finite’ languages. Such a set includes all finite languages and at least one infinite language. Intuitively, if the learner infers that the target language is an infinite language, whereas the actual target language is a finite language that is contained in the infinite language, then the learner will not encounter any contradicting evidence, and will never converge onto the correct language. This result holds in greatest generality: ‘algorithm’ here includes any function from text to language.

Probably almost correct

A common criticism of Gold’s framework is that the learner has to identify exactly the right language. For practical purposes, it might be sufficient that the learner acquires a grammar that is almost correct. Box 1 explains various extensions of the Gold framework, and in particular the approach of statistical learning theory. Here, the crucial requirement is that the learner converges with high probability to a language that is almost correct. Statistical learning theory also shows there is no procedure that can learn the set of all regular languages, thereby confirming the necessity of an innate UG. Some learning theories provide more information for the learner and thus allow larger classes of languages to be learnable, but no learning theory admits an unrestricted search space.

Learning finite languages

Some readers might think that the arguments of learning theory rely on subtle properties of infinite languages. Let us therefore consider finite languages. In the context of statistical learning theory, the set of all finite languages cannot be learned. In the Gold framework, the set of all finite languages can be learned, but only by memorization: the learner will identify the correct language only after having heard all sentences of this language. A learner that considers the set of all finite languages has no possibility for generalization: the learner can never extrapolate beyond the sentences he has already encountered. This is not the case for natural language acquisition: we can always say new sentences.

Let us consider a finite set of finite languages. Suppose there are 3 sentences, S_1, S_2, S_3 . Hence there are 8 possible languages. Suppose learner A considers all 8 languages, while learner B considers only 2 languages, for example $L_1 = \{S_1, S_2\}$ and $L_2 = \{S_3\}$. If learner A receives sentence S_1 , he has no information whether sentences S_2 or S_3 will be part of the target language or not. He can only identify the target language after having heard all sentences. If learner B receives sentence S_1 he knows that S_2 will be part of the language, whereas S_3

will not. He can extrapolate beyond his experience. The ability to search for underlying rules requires a restricted search space.

The necessity of innate expectations

We can now state in what sense there has to be an innate UG. The human brain is equipped with a learning algorithm, A_H , which enables us to learn certain languages. This algorithm can learn each of the existing 6,000 human languages and presumably many more, but it is impossible that A_H could learn every computable language. Hence, there is a restricted set of languages that can be learned by A_H . UG is the theory of this restricted set.

Learning theory suggests that a restricted search space has to exist before data. By 'data' we mean linguistic or other information the child uses to learn language or modify its language acquisition procedure. Therefore, in our terminology, 'before data' is equivalent to 'innate'. In this sense, learning theory shows there must be an innate UG, which is a consequence of the particular learning algorithm, A_H , used by humans. Discovering properties of A_H requires the empirical study of neurobiological and cognitive functions of the human brain involved in language acquisition. Some aspects of UG, however, might be unveiled by studying common features of existing human languages. This has been a major goal of linguistic research during the past decades. A particular approach is the 'principles and parameters theory', which assumes that the child comes equipped with innate principles and has to set parameters that are specific for individual languages^{38,39}. Another approach is 'optimality theory', where learning a specific language is ordering innate constraints⁴⁰.

There is some discourse as to whether the learning mechanism, A_H , is language specific or general purpose⁴¹. Ultimately this is a question about the particular architecture of the brain and which neurons participate in which computations, but one cannot deny that there is a learning mechanism, A_H , that operates on linguistic input and enables the child to learn the rules of human language. This mechanism can learn a restricted set of languages; the theory of this set is UG. The continuing debate around an innate UG should not be whether there is one, but what form it takes^{41–43}. One can

dispute individual linguistic universals^{44,45}, but one cannot generally deny their existence.

Neural networks are an important tool for modelling the neural mechanisms of language acquisition. The results of learning theory also apply to neural networks: no neural network can learn an unrestricted set of languages⁴⁶.

Sometimes it is claimed that the logical arguments for an innate UG rest on particular mathematical assumptions of generative grammars that deal only with syntax and not with semantics. Cognitive^{47,48} and functional linguistics⁴⁹ are not based on formal language theory, but use psychological objects such as symbols, categories, schemas and images. This does not remove the necessity of innate restrictions. The results of learning theory apply to any learning process, where a 'rule' has to be learned from some examples. Generalization is an inherent feature of any model of language acquisition, and applies to semantics, syntax and phonetics. Any procedure for successful generalization has to choose from a restricted range of hypotheses.

The results of learning theory also apply to learning mappings between linguistic form and meaning. If meaning is to be explicitly considered, then a language is not a set of sentences, but a set of sentence-meaning pairs (Box 2). The task of language acquisition is then to learn grammars that generate sentence-meaning pairs. Such grammars are also part of the Chomsky hierarchy, and there exists no learning procedure that can succeed on an unrestricted set of such languages.

What is special about language acquisition?

Usually when we learn the grammar of generative systems, such as chess or arithmetic, somebody tells us the rules. We do not have to guess the moves of chess by looking at chess games. In contrast, the process of language acquisition occurs without being instructed about rules; neither teachers nor learners are aware of the rules. This is an important difference: if the learner is told the grammar of a language, then the set of all computable languages is learnable by an algorithm that memorizes the rules.

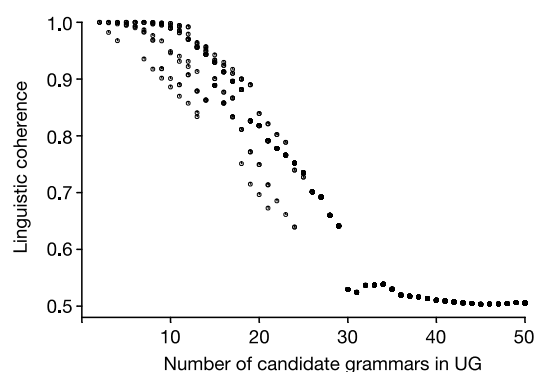


Figure 4 Linguistic coherence evolves if universal grammar (UG) is sufficiently specific. The figure shows equilibrium solutions of the language dynamical equation. Linguistic coherence, ϕ , is the probability that one individual says a sentence that is understood by another individual. UG specifies n candidate grammars. The similarity between any two candidate grammars, s_{ij} , is a random number from a uniform distribution on $[0, 1]$. The language acquisition device is a memoryless learner receiving $N = 100$ example sentences. For $n > 30$, all candidate grammars are represented in the population with similar frequencies; the linguistic coherence, ϕ , is about 1/2, which means complete randomness. For $n < 30$ the equilibrium is dominated by a single grammar. For each value of n there can be multiple equilibria dominated by different grammars. The equilibrium that is reached depends on the initial condition. Coherence is required for adaptation of language and selection of UG for linguistic function.

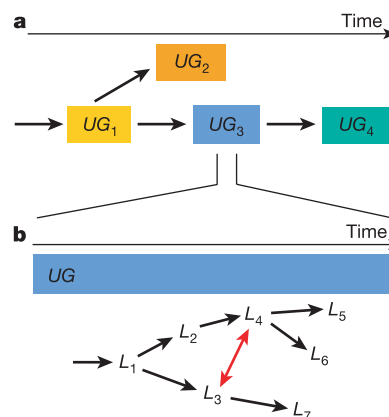


Figure 5 Two aspects of language evolution. **a**, There is a biological evolution of universal grammar (UG) via genetic modifications that affect the architecture of the human brain and the class of languages it can learn. UG can change as a consequence of (1) random variation (neutral evolution), (2) as a by-product of selection for other cognitive function or (3) under selection for language acquisition and communication. At some point in the evolutionary history of humans, a UG arose that allowed languages with infinite expressibility. **b**, On a faster timescale, there is cultural evolution of language constrained by a constant UG. Languages change by (1) random variation, (2) by contact with other languages (red arrow), (3) by hitch-hiking on other cultural inventions, or (4) by selection for increased learnability and communication. Although many language changes in historical linguistics might be neutral, a global picture of language evolution must include selection.

Evolutionary language theory

Humans and chimpanzees separated some 5 million years ago. Chimpanzees have a complex system of conceptual understanding and rich social interactions, but they do not have communication comparable to human language. The central question of the origin of human language is which genetic modifications led to changes in brain structures that were decisive for human language. Given the enormous complexity of this trait, we should expect several incremental steps guided by natural selection. In this process, evolution will have reused cognitive features that evolved long ago and for other purposes.

Understanding language evolution requires a theoretical framework explaining how darwinian dynamics lead to fundamental properties of human language such as arbitrary signs, lexicons, syntax and grammar^{50–62}. Here we outline a minimalist program combining formal language theory, learning theory and evolutionary theory.

The basic approach is similar to evolutionary game theory. There is a population of individuals. Each individual uses a particular language. Individuals talk to each other. Successful communication results in a pay-off that contributes to fitness. Offspring inherit—subject to mutations—a mechanism to learn language and a UG. They use this mechanism to learn—subject to mistakes—the language of their parents or others.

Cultural evolution of language with constant universal grammar

From a biological perspective, language is not the property of an individual, but the extended phenotype of a population. Let us consider a population where all individuals have the same UG, and let us assume that UG does not change from one generation to the next. Suppose that (in accordance with principles and parameters theory or optimality theory) UG specifies a finite number of languages $L_1, \dots, L_n$. Each individual in the population will speak one of those n languages.

We need a model of language that allows us to calculate the communicative pay-off, F_{ij} , for an individual using L_i talking to an individual using L_j . Box 2 outlines a fairly general approach, based on the assumption that a language can be seen as an infinite binary matrix linking phonetic forms to semantic forms. The languages, L_i , can differ in their intrinsic communicative pay-off, F_{ii} , which depends on ambiguity and expressive power. Some languages could be finite, others infinite.

Denote by x_i the relative abundance of speakers of L_i . The fitness of L_i is given by $f_i = \sum_{j=1}^n x_j F_{ij}$. Individuals leave offspring proportional to their pay-off. The probability that a child will develop L_j if the parent uses L_i is given by Q_{ij} . The ‘learning matrix’, Q , depends on the learning algorithm and UG. The language dynamical equation⁶² is given by:

$$\frac{dx_j}{dt} = \sum_{i=1}^n f_i(\mathbf{x}) Q_{ij} x_i - \phi(\mathbf{x}) x_j \quad j = 1, \dots, n \quad (1)$$

The term $-\phi(\mathbf{x}) x_j$ ensures that $\sum_i x_i = 1$. The variable $\phi(\mathbf{x}) = \sum_i f_i(\mathbf{x}) x_i$ denotes the average fitness of the population, and is a measure for linguistic coherence. The dynamics can also be interpreted in a purely cultural sense: individuals that communicate successfully are more likely to influence language acquisition of others. Equation (1) describes selection of languages for increased communicative function, F_{ii} , and increased learnability, Q_{ij} .

For low accuracy of language acquisition, when Q is far from the identity matrix, there is no predominating language in the population, and the linguistic coherence is low. As the accuracy of language acquisition increases, and Q gets closer to the identity matrix, equilibrium solutions arise where a particular language is more abundant than others. The population has achieved linguistic coherence. The ‘coherence threshold’ specifies the minimum specificity of UG that is required for linguistic coherence (Fig. 4).

For certain learning mechanisms we can calculate the coherence threshold. The ‘memoryless learner’ starts with a randomly chosen language and stays with it as long as the input sentences are compatible with this language. If a sentence arrives that is not compatible, then the learner picks at random another language. The learner does not memorize which languages have already been rejected. The process stops after N sentences. Another mechanism is the ‘batch learner’, which memorizes N sentences, and at the end chooses the language that is most compatible with all N sentences.

If the similarity coefficients between languages, s_{ij} (Box 2), are constant, $s_{ij} = s$ and $s_{ii} = 1$, then the memoryless learner has a coherence threshold $N > C_1 n$, whereas the batch learner has a coherence threshold $N > C_2 \log n$. If the s_{ij} values are taken from a uniform random distribution on the interval $[0, 1]$ and if $s_{ii} = 1$, then the memoryless learner has a coherence threshold $N > C_3 n \log n$, whereas the batch learner has a coherence threshold $N > C_4 n$ (refs 63, 64). C_1 to C_4 are some constants. These conditions provide boundaries for the actual learning mechanism used by humans, which is arguably better than the memoryless learner and worse than the batch learner. The coherence threshold relates a life-history parameter of humans, N , to the maximum size of the search space, n , of UG.

Evolution of universal grammar

Evolution of UG requires variation of UG. (UG is in fact neither a grammar nor universal.) Imagine a population of individuals using UGs U_1 to U_M . Each U_I admits a subset of n grammars and determines a particular learning matrix Q^I . U_I mutates genetically to U_J with probability W_{IJ} . Deterministic population dynamics are given by:

$$\frac{dx_{Jj}}{dt} = \sum_{I=1}^M W_{IJ} \sum_{i=1}^n f_{Ii} Q_{ij}^I x_{Ii} - \phi x_{Jj} \quad j = 1, \dots, n \quad J = 1, \dots, M \quad (2)$$

This equation describes mutation and selection among M different universal grammars. The relative abundance of individuals with UG U_J speaking language L_j is given by x_{Jj} . At present little is known about the behaviour of this system. In the limit of no mutation among UGs, $W_{II} = 1$, we find that the selective dynamics often lead to the elimination of all but one UG, but sometimes coexistence of different UGs can be observed. Equation (2) describes two processes on different timescales: the biological evolution of UG and the cultural evolution of language (Fig. 5).

The ability to induce a coherent language is a major selective criterion for UG. A UG that induces linguistic coherence allows language adaptation and can be selected for linguistic function. There is also a trade-off between learnability and adaptability: a small search space (small n) is more likely to lead to linguistic coherence, but might exclude languages with high communicative pay-off.

Because the necessity of a restricted search space applies to any learning task, we can use an extended concept of UG for animal communication. During primate evolution, there was a succession of UGs that finally led to the UG of currently living humans. At some point a UG emerged that allowed languages of unlimited expressibility. Such evolutionary dynamics are described by equation (2).

Historical linguistics

The language dynamical equation (1) can be used to study language change in the context of historical linguistics^{65–68}. Languages change because the transmission from one generation to the next is not perfect. UG limits the type of variation that can occur. In the context of the principles-and-parameters theory, changes in syntax arise because children acquire different parameter settings⁶⁶. Grammaticalization⁶⁹ is the process where lexical items take on grammatical function. Creolization is the formation of a new language by

children receiving mixed input^{70,71}. All such language changes can be studied mathematically. Many language changes are selectively neutral. Hence, we can use a neutral version of our approach possibly in conjunction with small population sizes and stochastic and spatial population dynamics. These are open problems.

Outlook

We have reviewed mathematical descriptions of language on three different levels: formal language theory, learning theory and evolution. These approaches need to be combined: ideas of language should be discussed in the context of acquisition, and ideas of acquisition in the context of evolution.

Some theoretical questions are: what is the interplay between the biological evolution of UG and the cultural evolution of language? What is the mechanism for adaptation among the various languages generated by a given UG? In terms of the principles-and-parameters theory, can we estimate the maximum number of parameters compatible with the coherence threshold? Some empirical questions are: what is the actual language learning algorithm used by humans? What are the restrictions imposed by UG? Can we identify genes that are crucial for linguistic or other cognitive functions? What can we say about the evolution of those genes?

The study of language as a biological phenomenon will bring together people from many disciplines including linguistics, cognitive science, psychology, genetics, animal behaviour, evolutionary biology, neurobiology and computer science. Fortunately we have language to talk to each other. □

doi:10.1038/nature00771.

1. Pinker, S. & Bloom, A. Natural language and natural selection. *Behav. Brain Sci.* **13**, 707–784 (1990).
2. Jackendoff, R. Possible stages in the evolution of the language capacity. *Trends Cogn. Sci.* **3**, 272–279 (1999).
3. Bickerton, D. *Language and Species* (Univ. Chicago Press, Chicago, 1990).
4. Lightfoot, D. *The Development of Language: Acquisition, Changes and Evolution* (Blackwell, Oxford, 1999).
5. Brandon, R. & Hornstein, N. From icon to symbol: Some speculations on the evolution of natural language. *Phil. Biol.* **1**, 169–189 (1986).
6. Hurford, J. R., Studdert-Kennedy, M. A. & Knight, C. (eds) *Approaches to the Evolution of Language* (Cambridge Univ. Press, Cambridge, UK, 1998).
7. Newmeyer, F. Functional explanation in linguistics and the origins of language. *Lang. Commun.* **11**, 3–28, 97–108 (1991).
8. Lieberman, P. *The Biology and Evolution of Language* (Harvard Univ. Press, Cambridge, Massachusetts, 1984).
9. Maynard Smith, J. & Szathmari, E. *The Major Transitions in Evolution* (Freeman Spektrum, Oxford, 1995).
10. Hawkins, J. A. & Gell-Mann, M. *The Evolution of Human Languages* (Addison-Wesley, Reading, Massachusetts, 1992).
11. Aitchinson, J. *The Seeds of Speech* (Cambridge Univ. Press, Cambridge, UK, 1996).
12. Cavalli-Sforza, L. L. Genes, peoples and languages. *Proc. Natl Acad. Sci. USA* **94**, 7719–7724 (1997).
13. Gopnik, M. & Crago, M. Familial aggregation of a developmental language disorder. *Cognition* **39**, 1–50 (1991).
14. Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. A forhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001).
15. Deacon, T. *The Symbolic Species* (Penguin, London, 1997).
16. Vargha-Khadem, F. et al. Neural basis of an inherited speech and language disorder. *Proc. Natl Acad. Sci. USA* **95**, 12695–12700 (1998).
17. Smith, W. J. *The Behaviour of Communicating* (Harvard Univ. Press, Cambridge, UK, 1977).
18. Dunbar, R. *Grooming, Gossip, and the Evolution of Language* (Cambridge Univ. Press, Cambridge, UK, 1996).
19. Fitch, W. T. The evolution of speech: a comparative review. *Trends Cogn. Sci.* **4**, 258–267 (2000).
20. Hauser, M. D. *The Evolution of Communication* (Harvard Univ. Press, Cambridge, Massachusetts, 1996).
21. Chomsky, N. A. *Syntactic Structures* (Mouton, New York, 1957).
22. Harrison, M. A. *Introduction to Formal Language Theory* (Addison-Wesley, Reading, Massachusetts, 1978).
23. Gold, E. M. Language identification in the limit. *Informat. Control* **10**, 447–474 (1967).
24. Vapnik, V. N. & Chervonenkis, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Prob. Appl.* **17**, 264–280 (1971).
25. Valiant, L. G. A theory of learnable. *Commun. ACM* **27**, 436–445 (1984).
26. Vapnik, V. N. *Statistical Learning Theory* (Wiley, New York, 1998).
27. Osherson, D., Stob, M. & Weinstein, S. *Systems That Learn* (MIT Press, Cambridge, Massachusetts, 1986).
28. Pinker, S. Formal models of language learning. *Cognition* **7**, 217–283 (1979).
29. Pullum, G. K. & Gazdar, G. Natural languages and context free languages. *Linguist. Phil.* **4**, 471–504 (1982).
30. Shieber, S. M. Evidence against the context-freeness of natural language. *Linguist. Phil.* **8**, 333–343 (1985).
31. Chomsky, N. A. *Lectures on Government and Binding: The Pisa Lectures* (Foris, Dordrecht, 1984).
32. Sadock, J. M. *Autolexical Syntax: A Theory of Parallel Grammatical Representations. Studies in Contemporary Linguistics* (Univ. Chicago Press, Chicago, 1991).
33. Bresnan, J. *Lexical-Functional Syntax* (Blackwells, London, 2001).
34. Pollard, C. J. & Sag, I. A. *Head-Driven Phrase Structure Grammar* (Univ. Chicago Press, Chicago, 1994).
35. Chomsky, N. *Language and Mind* (Harcourt Brace Jovanovich, New York, 1972).
36. Wexler, K. & Culicover, P. *Formal Principles of Language Acquisition* (MIT Press, Cambridge, Massachusetts, 1980).
37. Jackendoff, R. *Foundations of Language* (Oxford Univ. Press, Oxford, 2001).
38. Chomsky, N. in *Explanation in Linguistics* (eds Hornstein, N. & Lightfoot, D.) 123–146 (Longman, London, 1981).
39. Baker, M. C. *Atoms of Language* (Basic Books, New York, 2001).
40. Prince, A. & Smolensky, P. Optimality: From neural networks to universal grammar. *Science* **275**, 1604–1610 (1997).
41. Elman, J. L. *Rethinking Innateness* (MIT Press, Cambridge, Massachusetts, 1996).
42. Tomasello, M. *The Cultural Origins of Human Cognition* (Harvard Univ. Press, Cambridge, Massachusetts, 1999).
43. Sampson, G. *Educating Eve: The Language Instinct Debate* (Cassell Academic, London, 1999).
44. Greenberg, J. H., Ferguson, C. A. & Moravcsik, E. A. (eds) *Universals of Human Language* (Stanford Univ. Press, Stanford, 1978).
45. Comrie, B. *Language Universals and Linguistic Typology* (Univ. Chicago Press, Chicago, 1981).
46. Geman, S., Bienenstock, E. & Doursat, R. Neural networks and the bias/variance dilemma. *Neural Comput.* **4**, 1–58 (1992).
47. Langacker, R. *Foundations of Cognitive Linguistics* Vol. 1 (Stanford Univ. Press, Stanford, 1987).
48. Lakoff, G. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind* (Univ. Chicago Press, Chicago, 1987).
49. Bates, E. & MacWhinney, B. *Language Acquisition: The State of the Art* (Cambridge Univ. Press, Cambridge, 1982).
50. Aoki, K. & Feldman, M. W. Toward a theory for the evolution of cultural communication: Coevolution of signal transmission and reception. *Proc. Natl Acad. Sci. USA* **84**, 7164–7168 (1987).
51. Hurford, J. R. Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua* **77**, 187–222 (1989).
52. Cangelosi, A. & Parisi, D. *Simulating the Evolution of Language* (Springer, London, 2002).
53. Kirby, S. & Hurford, J. *Proc. Fourth European Conf. on Artificial Life* (eds Husbands, P. & Harvey, I.) 493–502 (MIT Press, Cambridge, Massachusetts, 1997).
54. Steels, L. *Proc. Fifth Artificial Life Conf.* (eds Langton, C. G. & Shimohara, T.) 113–131 (MIT Press, Tokyo, 1996).
55. Nowak, M. A. & Krakauer, D. C. The evolution of language. *Proc. Natl Acad. Sci. USA* **96**, 8028–8033 (1999).
56. Nowak, M. A., Plotkin, J. B. & Jansen, V. A. A. Evolution of syntactic communication. *Nature* **404**, 495–498 (2000).
57. Komarova, N. L. & Nowak, M. A. Evolutionary dynamics of the lexical matrix. *Bull. Math. Biol.* **63**, 451–485 (2001).
58. Christiansen, M. H., Dale, R. A. C., Ellefson, M. R. & Conway, C. M. in *Simulating the Evolution of Language* (eds Cangelosi, A. & Parisi, D.) 165–187 (Springer, London, 2002).
59. Hashimoto, T. & Ikegami, T. Emergence of net-grammar in communicating agents. *Biosystems* **38**, 1–14 (1996).
60. Hazlehurst, B. & Hutchins, E. The emergence of propositions from the coordination of talk and action in a shared worlds. *Lang. Cogn. Process.* **13**, 373–424 (1998).
61. Pinker, S. *The Language Instinct* (Morrow, New York, 1994).
62. Nowak, M. A., Komarova, N. L. & Niyogi, P. Evolution of universal grammar. *Science* **291**, 114–118 (2001).
63. Komarova, N. L. & Rivin, I. Mathematics of learning. Preprint math.PR/0105235 at (<http://lanl.arXiv.org>) (2001).
64. Rivin, I. Yet another zeta function and learning. Preprint cs.LG/0107033 at (<http://lanl.arXiv.org>) (2001).
65. Lightfoot, D. *How to Set Parameters: Arguments from Language Change* (MIT Press, Cambridge, Massachusetts, 1991).
66. Kroch, A. Reflexes of grammar in patterns of language change. *Lang. Variat. Change* **1**, 199–244 (1989).
67. Wang, W. S. Y. in *The Origins and Past of Modern Humans* (eds Omoto, K. & Tobias, P. V.) 247–262 (World Scientific, Singapore, 1998).
68. Niyogi, P. & Berwick, R. C. Evolutionary consequences of language learning. *Linguist. Phil.* **20**, 697–719 (1997).
69. Hopper, P. & Traugott, E. *Grammaticalization* (Cambridge Univ. Press, Cambridge, 1993).
70. de Graff, M. *Language Creation and Language Change: Creolization, Diachrony and Development* (MIT Press, Cambridge, MA, 1999).
71. Mufwene, S. *The Ecology of Language Evolution* (Cambridge Univ. Press, Cambridge, 2001).
72. Angluin, D. Learning regular sets from queries and counterexamples. *Informat. Comput.* **75**, 87–106 (1987).
73. Angluin, D. & Kharitonov, M. When won't membership queries help? *J. Comput. Syst. Sci.* **50**, 336–355 (1995).
74. Gasarch, W. & Smith, C. Learning via queries. *J. Assoc. Comput. Machin.* **39**, 649–674 (1992).

Acknowledgements

Support from the David and Lucille Packard foundation, the Leon Levy and Shelby White initiatives fund, the Florence Gould foundation, the Ambrose Monell foundation, the National Science Foundation and J. E. Epstein is acknowledged.

Correspondence and requests for materials should be addressed to M.A.N. (e-mail: nowak@ias.edu).

Hsp90 as a capacitor of phenotypic variation

Christine Queitsch*, Todd A. Sangster† & Susan Lindquist*‡

* Department of Molecular Genetics and Cell Biology, and † Committee on Genetics, Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637, USA

Heat-shock protein 90 (Hsp90) chaperones the maturation of many regulatory proteins and, in the fruitfly *Drosophila melanogaster*, buffers genetic variation in morphogenetic pathways. Levels and patterns of genetic variation differ greatly between obligatorily outbreeding species such as fruitflies and self-fertilizing species such as the plant *Arabidopsis thaliana*. Also, plant development is more plastic, being coupled to environmental cues. Here we report that, in *Arabidopsis* accessions and recombinant inbred lines, reducing Hsp90 function produces an array of morphological phenotypes, which are dependent on underlying genetic variation. The strength and breadth of Hsp90's effects on the buffering and release of genetic variation suggests it may have an impact on evolutionary processes. We also show that Hsp90 influences morphogenetic responses to environmental cues and buffers normal development from destabilizing effects of stochastic processes. Manipulating Hsp90's buffering capacity offers a tool for harnessing cryptic genetic variation and for elucidating the interplay between genotypes, environments and stochastic events in the determination of phenotype.

During evolution the remodelling of shapes and functions draws on genetic differences between individuals and populations. However, species phenotypes must generally be robust to genetic variation and environmental change, requiring buffering systems to ensure normal development ('canalization')^{1,2}. Environmental stress can reveal genetic variants, presumably because it compromises buffering systems. If selected for, these uncovered phenotypes can lead to heritable changes in plants and animals (assimilation)^{3–5}. Recent work provided evidence that the chaperone Hsp90 can serve as such a buffer in *Drosophila melanogaster*. Remarkably, it can do so in a multitude of different morphological pathways⁶.

In all eukaryotes tested, Hsp90 is essential, abundant at normal temperatures, and induced by stress. Under physiological conditions, Hsp90 dynamically interacts with a diverse but highly select set of inherently unstable 'client' proteins (for example, kinases and transcription factors). Typically, it keeps these metastable proteins poised for activation until they are stabilized by conformational changes, such as those associated with signal transduction. The requirement of many principal regulatory proteins for Hsp90 renders entire pathways sensitive to decreases in its function^{7–9}.

In *Drosophila*, challenging Hsp90 function by mutation, pharmacological inhibition or environmental stress can produce a profusion of morphological changes affecting virtually every structure of the fly. Notably, the particular change observed in an individual fly depends on previously silent genetic variation. In the two cases tested, multiple polymorphisms affecting specific developmental pathways could be enriched by selection so that the traits were expressed even after Hsp90 function was restored. Thus, it appears that Hsp90 allows the storage and release of genetic variation in *Drosophila*⁶ as a consequence of its essential function in chaperoning regulators of growth and development. If so, this effect might be conserved in other organisms, potentially influencing the pace and nature of evolution.

The sessile, photosynthetic plant *Arabidopsis thaliana* has a vastly different lifestyle from the vagile, heterotrophic *Drosophila* and is evolutionarily distant. Moreover, the pattern and level of genetic variation differ profoundly between this inbreeding species and obligatorily outcrossing species such as the fly. Although poly-

morphisms are present within *Arabidopsis* populations and outcrossing does occur (estimated frequency ~0.3%; ref. 10), heterozygosity per individual is extremely low^{10,11}. Because heterozygosity is such an important means of maintaining genetic variability and phenotypic robustness, it is difficult to predict whether a buffering system like Hsp90 would act in the accumulation and release of variation in such organisms.

Additional features of *Arabidopsis* provide a singular opportunity to ask whether Hsp90 influences other global aspects in the manifestation of phenotype. The *Arabidopsis* research community has gathered specimens (accessions) from diverse regions of the globe, representing virtually homozygous, divergent genomes. These accessions can be crossed to produce large numbers of virtually identical, highly heterozygous F₁ progeny, as well as recombinant inbred lines representing homozygous genetic mosaics of parental genomes. Furthermore, in *Arabidopsis* (as in plants generally), development is highly plastic: morphologies vary dramatically with even moderate changes in environment (for example, light intensity)^{12–15}. These circumstances provide an opportunity to address two important questions not yet addressed in any system. First, does Hsp90 have a role in linking environmental change to the developmental dynamics that govern morphogenesis in an individual organism's lifetime? And, if so, can we assess how the strength of this linkage varies with genotype? Second, does Hsp90 influence the potentially positive and negative effects of heterozygosity on developmental stability?

Pharmacological inhibition of Hsp90 in seedlings

The *Arabidopsis* genome contains seven Hsp90 genes^{16,17}. Therefore, we used pharmacological inhibition to reduce Hsp90 function. The drugs geldanamycin (GDA) and radicicol are structurally unrelated; however, each inhibits Hsp90 function through multiple, distinct contacts with residues in its highly unusual, evolutionarily conserved nucleotide-binding pocket^{18,19}. Drug-binding residues are poorly conserved in the other proteins with this unconventional ATP-binding fold. An additional advantage is that both drugs are inactivated by prolonged exposure to light, thereby allowing the recovery and propagation of plants with early abnormal phenotypes caused by Hsp90 inhibition.

In the two accessions (ecotypes) initially examined, Columbia (Col) and Landsberg *erecta* (Ler), each drug produced dosage-

‡ Present address: Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

dependent phenotypes at a concentration that correlated with its affinity for Hsp90. For radicicol (dissociation constant K_d for binding Hsp90 is 19 nM), phenotypes appeared between 2 and 50 nM; for GDA ($K_d = 1.2 \mu\text{M}$)¹⁸, phenotypes appeared between 0.15 and 5 μM . Above the K_d , both drugs caused multiple phenotypes and reduced viability. At slightly lower concentrations, most plants remained healthy and unaffected but some exhibited strongly altered phenotypes. For example, at 1.0 μM GDA, 5–8% of seedlings showed strong morphological abnormalities, affecting shape, colour and expansion of cotyledons; shape, colour and presence of true leaves; shape and length of hypocotyls; root morphology; and the orientation of rosettes, roots or whole seedlings (Fig. 1 and data not shown). In most cases, phenotypes were distinct and not accompanied by a general failure to thrive. Without the drug, only 1–2% of plants showed far more subtle variant morphologies. Most importantly, the two chemically unrelated drugs, GDA (a benzoquinone ansamycin) and radicicol (a macrolactone), produced the same spectrum of phenotypes.

Next, we tested several geographically and environmentally diverse *Arabidopsis* accessions (see Methods). Again, most plants bearing abnormal phenotypes on GDA had an otherwise healthy appearance. Particular phenotypes were not restricted to particular accessions, but the frequencies of their appearance varied reproducibly. For example, rare plants of all accessions exhibited abnormalities in true leaves, including radially symmetrical leaves, deformed shapes, and missing leaves, but in the Shadara accession about 30% of seedlings had these phenotypes (Fig. 1a). Shadara also

frequently produced distorted rosettes with juxtaposed cotyledons (Fig. 1b). Dwarfed plants with cotyledons that accumulated purple pigment were most common in Col (Fig. 1c). Curled hypocotyls were most frequent in Ler (Fig. 1d). For comparison, a *Ler* seedling grown without GDA is shown in Fig. 1e.

Phenotypic variation specific to recombinant inbred lines

The predisposition of different accessions to different phenotypes suggested a genetic contribution to the phenotypic variation buffered by Hsp90. To test this, we examined recombinant inbred lines (RI lines) originating from crosses between two accessions followed by single-seed self-propagation for eight generations. Each line should be homozygous at almost all loci, with different lines representing various mosaics of the parental genomes²⁰. If a decrease in Hsp90 function uncovers new phenotypes due to genetic variation, GDA-induced phenotypes should vary between RI lines but tend to be shared within them. Here, we present work with 50 RI lines derived from a cross between Cape Verde Island (Cvi) and *Ler* (Cvi/*Ler*, CS22477 base set, Arabidopsis Biological Resource Center, ABRC) accessions. Col/*Ler* RI lines behaved similarly, although phenotypic variation was not as dramatic (data not shown), probably because these accessions are more closely related^{20,21}.

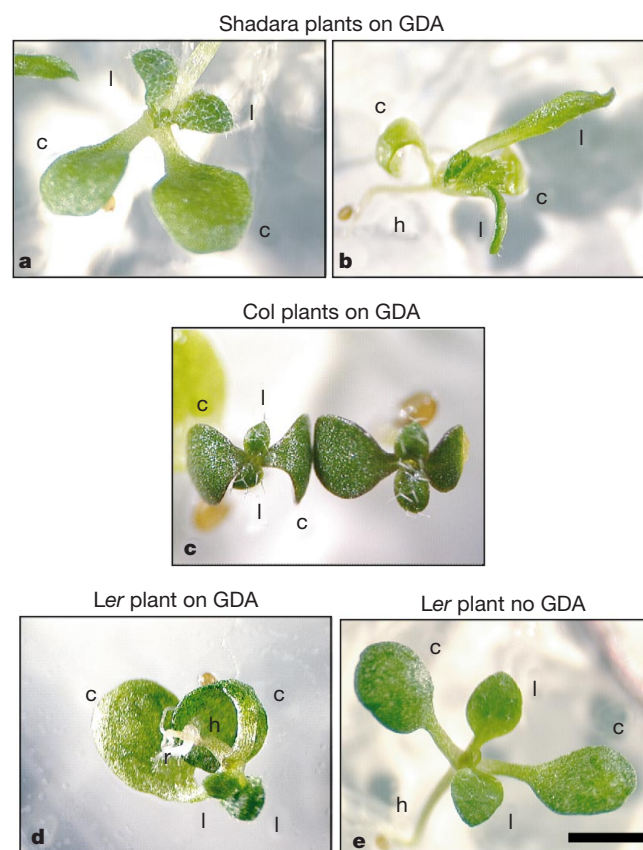


Figure 1 Common GDA-dependent phenotypes in Shadara, Col and *Ler* accessions. **a–d**, Seedlings grown with 1 μM GDA. **e**, *Ler* seedling grown without GDA. Note typical symmetry of *Arabidopsis* rosette and oval-shaped, horizontally oriented, flat leaves. **a**, Shadara, juxtaposed cotyledons. **b**, Shadara, deformed and radially symmetrical true leaves. **c**, Col, dwarfed with dark green, epinastic cotyledons and true leaves. **d**, *Ler*, extremely curled hypocotyl with roots partially extended in the air. c, cotyledons; l, true leaves; h, hypocotyl; r, root. Scale bar, 2 mm.

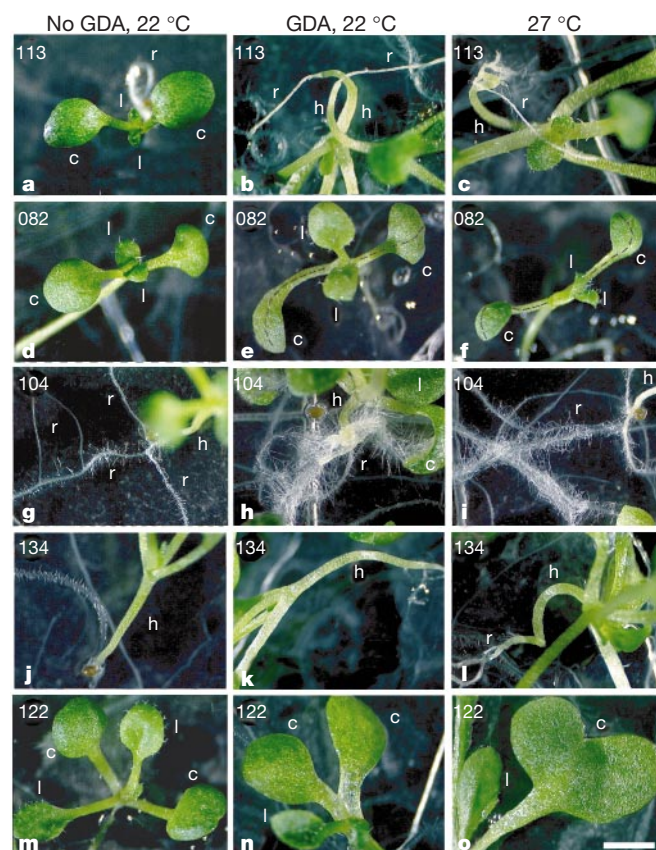


Figure 2 The same RI line-specific phenotypes are uncovered when buffering capacity is challenged by GDA (middle) and by growth at 27 °C (right). **a, d, g, j, m**, RI line seedlings grown without GDA. **b, c**, Line 113 seedlings, extreme hypocotyl curls and roots partially extended into air. **e, f**, Line 082 seedlings, S-shaped rosettes with vertically oriented leaf blades. **h, i**, Line 104 seedlings, abundant root hair growth. **k, l**, Line 134 seedlings, bent hypocotyls with rosette touching the medium surface. **n, o**, Line 122 seedlings, juxtaposed cotyledons (similar to those in Shadara) fused to varying degrees. In extreme cases, seedlings appeared to have only one very large, malformed cotyledon and one apparently healthy true leaf. Because GDA activity is lost after prolonged light exposure (see Methods), this phenotype (and other leaf phenotypes) did not appear in later-developing leaves. Plant organs are indicated, as defined in Fig. 1. Scale bar: 2 mm for **a–f, j–m**; 1 mm for **g–i, n, o**.

The most striking Hsp90-buffered phenotypes in *Cvi/Ler* RI lines were distinct and restricted to particular lines (Fig. 2, middle; Table 1). For example, seedlings of line CS22113 (113) developed normally for about 8 days. Thereafter, their hypocotyls started to curl, lifting the root above the medium surface (Fig. 2b). By day 10, about 30% of the seedlings showed extreme hypocotyl curls and unusually large, round cotyledons. For seedlings of line 082, leaf blades of cotyledons were positioned vertically instead of horizontally, twisting the seedling rosette in an S shape (Fig. 2e). Seedlings of lines 104 (Fig. 2h) and 159 (data not shown) displayed a profuse increase in root-hair growth. Some line 104 seedlings also accumulated purple pigment in cotyledons and emerging adult leaves (data not shown). For line 134 seedlings, arched hypocotyls were observed (Fig. 2k).

Other RI lines exhibited phenotypes that were variable but appeared related: fused cotyledons and distorted rosettes (line 122, Fig. 2n), radially symmetrical leaves and missing true leaves (lines 119 and 120, data not shown), and other leaf deformities (line 119, data not shown). Lines 119 and 120 also showed occasional trichomes on the abaxial side (underside) of the first true leaf (data not shown).

Notably, a few seedlings in most lines exhibited altered phenotypes without GDA. Owing to potential pitfalls of subjective analysis, all seedlings with detectable phenotypes were scored in Table 1. As with accessions (Fig. 1), these phenotypes were not only less frequent without GDA, but they were also generally more subtle in character (data not shown). They were, however, specific to individual lines and related to the stronger phenotypes observed with GDA. That is, although seedlings of line 113 sometimes exhibited slightly curled hypocotyls without GDA, none exhibited S-shaped rosettes, malformed true leaves, or abnormally abundant root-hair growth. Thus, GDA phenotypes were based on uncovering the genetic predispositions of different RI lines to produce different traits (Table 1), rather than to spurious drug effects. Some of the traits (for example, juxtaposed cotyledons) observed in RI lines were never observed in their parents, presumably because the lines represent new combinations of the original genotypes. These combinations of parental genomes produce the potential for a wide variety of genetically determined traits, which are expressed when Hsp90 chaperone capacity falls below a certain level.

The diversity and specificity of the phenotypes observed indicates that the level of Hsp90 inhibition required to uncover them is relatively modest. If inhibition had been strong enough to interfere with developmental pathways that have an absolute dependence on Hsp90 in this species, all plants would exhibit similar phenotypes. Moreover, the robust health of most seedlings displaying strong phenotypes suggested that although inhibition was sufficient to uncover certain genetic predispositions, it was not so severe as to

interfere with the general 'housekeeping' functions of Hsp90. To test whether our conditions induced a general stress response, eight randomly selected lines, including two that produced the strongest phenotypes (113 and 119), were examined for expression of Hsp101 and Hsp70 by western blotting. All exhibited normal Hsp inductions with a brief 38 °C heat shock²². None showed detectable inductions with GDA at their normal growth temperature (22 °C; data not shown). Certainly, phenotypes might be influenced by other Hsps, but the level of Hsp90 inhibition required to uncover the specific polymorphisms that can give rise to strong phenotypes was generally lower than that required to produce a detectable stress response.

Phenotypes uncovered by environmental change

If Hsp90-buffered variation acts in the storage and release of genetic variation, then moderate environmental changes might be sufficient to cross the threshold to express particular traits in plants carrying particular polymorphisms. To test this, we grew *Cvi/Ler* RI line seedlings without GDA, but at 27 °C instead of 22 °C. (As with GDA, this moderate temperature increase did not detectably induce other Hsps; data not shown.) All but one of the lines that exhibited strong phenotypes with GDA showed the same phenotypes at 27 °C (compare Fig. 2b with c, e with f, h with i, k with l, and n with o; Table 1). Even in the exceptional case, line 104, one of its two GDA-induced phenotypes (abundant root-hair growth) was expressed (Fig. 2i, Table 1). Furthermore, the ability of environmental change to reveal previously hidden genetic variation was precise: among the seven lines presented in Table 1, the phenotypes were distinct, whereas in the other 42 lines examined, some seedlings exhibited altered phenotypes at 27 °C, but rarely these. Other environmental conditions, such as the nature of the growth substratum, also had the capacity to uncover Hsp90-buffered traits, and these could function in a combinatorial fashion (see Supplementary Information).

Hsp90-dependent phenotypic plasticity

Next, we asked whether developmental plasticity was dependent on Hsp90 function, employing traits that could be scored objectively. The dark response is a classic example of developmental plasticity: in contrast to light-grown *Arabidopsis* seedlings, dark-grown seedlings have elongated hypocotyls; small, yellow and non-expanded cotyledons; and short roots^{23,24}. In Fig. 3, RI lines and their parents (red labels) are sorted by their degree of hypocotyl elongation when grown in the dark without GDA. Notably, the breadth of RI line responses was greater than those of parental lines (transgression). Thus, genetic variants affecting elongation were contributed by both parents and segregated in the original cross.

Table 1 Phenotypes specific to RI lines

RI line	Without GDA, 22 °C		With GDA, 22 °C		Without GDA, 27 °C	
	Phenotype*	Frequency†	Phenotype	Frequency	Phenotype	Frequency
CS22113	1	3.5% (1/28)	1	32% (9/28)	1	32% (9/28)
CS22082	2	1.7% (2/112)	2	32% (64/201)	2	95% (53/56)
CS22104	3	0% (0/75)	3	34% (29/84)	3	0% (0/47)
	4	16% (12/75)	4	52.4% (44/84)	4	38% (18/47)
CS22159	4	6.3% (8/127)	4	50.5% (53/105)	4	28% (18/63)
CS22134	5	0% (0/28)	5	87.5% (49/56)	5	46% (19/41)
CS22119	3	0% (0/28)	3	7% (2/28)	3	0% (0/28)
	6	7% (2/28)	6	68% (19/28)	6 or 7	90% (25/28)
	7	0% (0/28)	7	7% (2/28)		
CS22120	6	7% (2/28)	6	64% (18/29)	6	82% (23/28)
CS22122	6	0% (0/110)	6	2.5% (4/158)	6	0% (0/130)
	7	15% (17/110)	7	28% (44/158)	7	31.5% (41/130)

Phenotypes specific to RI lines appeared when buffering capacity was reduced by GDA or by a moderate temperature increase. These phenotypes were distinct and restricted to particular lines. In the absence of GDA at 22 °C, rare seedlings of some lines (113, 082, 119 and 120) exhibited phenotypes that were similar to those observed in the presence of GDA, but these were much more subtle. For lines 122, 104 and 159, phenotypes in the absence of GDA ranged from subtle to as strong as those observed with GDA.

* Phenotypes: 1, extreme hypocotyl curl and roots partially extended into air; 2, S-shaped rosettes with vertically oriented leaf blades; 3, accumulation of purple pigment; 4, abundant root hairs; 5, bent hypocotyls with rosette touching the medium surface; 6, malformed true leaves; 7, juxtaposed cotyledons, malformed cotyledons.

† Per cent penetrance, followed by affected plants/total number of plants scored (in parentheses).

The effects of GDA on hypocotyl elongation were uniform within lines but differed markedly between lines. In some (for example, lines 077 and 129; Fig. 3, thin arrows), the drug had little effect. In others (for example, lines 086 and 116; Fig. 3, thick arrows), it reduced elongation to nearly the values obtained in white light. GDA did not simply intensify the variation observed without the drug. Lines with naturally long or short hypocotyls (Fig. 3, grey bars) were found both in the group that was strongly affected by Hsp90 inhibition and in the group that was little affected. Both parental lines showed moderate reductions in hypocotyl length. Thus, the degree to which the dark response was affected segregated in RI lines, producing lines with extreme Hsp90 dependence and others with much lower dependence.

We examined the Hsp90 dependence of several other developmental plasticity traits: root growth in the dark promoted by sucrose in the medium; germination in the dark; the ability of cotyledons to become green after growth in the dark; and the ability of roots to respond to a change in the direction of gravity (Fig. 3, more details are provided in Supplementary Information). If Hsp90 buffers polymorphisms in many different environmental response pathways, we would expect the Hsp90 dependence of different traits to vary both within a given RI line and between RI lines. Indeed, in the same seedlings, measured on the same day, GDA affected different traits in different ways. For example, consider lines 132

and 079 (Fig. 3, blue labels): hypocotyl elongation and 'greening' were more strongly affected in 132; root elongation and germination were more affected in 079. Gravitropism responses were assessed in a separate experiment by rotating plants grown on vertical plates by 90°. This trait also showed different sensitivities to GDA in different lines, including one (129) in which roots turned more efficiently with GDA than without (Wilcoxon two-sample test, $P \leq 0.018$). Most notably, GDA's effects on specific plasticity responses seemed largely independent of each other (Spearman's rank tests, turning versus root elongation, $P \leq 0.7$; versus hypocotyl elongation, $P \leq 0.3$; versus germination rate, $P \leq 0.3$; germination rate versus root elongation, $P \leq 0.6$; versus hypocotyl elongation, $P \leq 0.5$).

Thus, the Hsp90-dependent variation was due neither to random nor to line-specific differences in drug uptake or toxicity. Rather, it resulted from genetic variation in different environmental response pathways that rendered them more or less dependent on full Hsp90 function. By QTL mapping, the uncovered phenotypic variation in hypocotyl elongation did not map to the cytoplasmic Hsp90 gene cluster (C.Q., J. Borevitz, J. Maloof, D. Weigel, J. Chory and S.L., unpublished observations). This confirms that, for this trait, the phenotypic variation uncovered by Hsp90 inhibition is not due to polymorphisms in Hsp90 itself but to a wide variety of polymorphisms elsewhere in the genome.

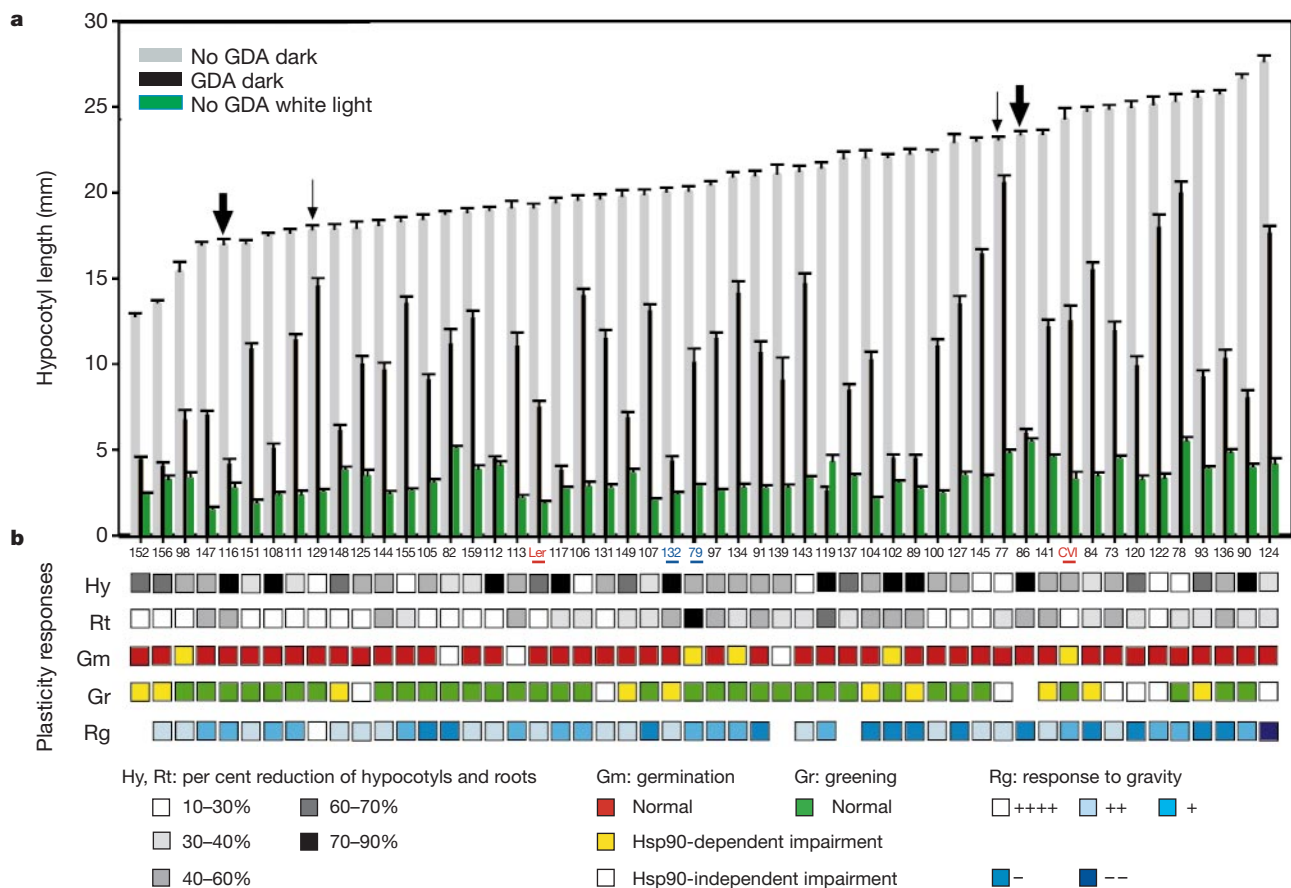


Figure 3 The effect of GDA on developmental plasticity responses differs between 50 Cvi/Ler RI lines and their parents. **a**, Response of hypocotyls to growth in the dark without GDA (grey bars), with GDA (black bars) and in white light (green bars). Error bars represent s.e.m. Red labels indicate parental accessions; blue labels indicate examples of RI lines for which GDA's effects on developmental responses differed. Thick arrows, examples of lines in which GDA reduced hypocotyl elongation to values close to those of seedlings grown in white light; thin arrows, lines for which hypocotyl length was little affected by GDA. **b**, Hy and Rt, degree of hypocotyl and root reduction: white, least affected lines;

black, most affected lines. Gm, germination scored by radicle emergence: red, nearly complete germination; yellow, poor germination only on GDA; white, poor germination without GDA and further reduction with GDA. Gr, greening: green, complete greening; yellow, greening partially or completely impaired only on GDA; white, greening impaired with or without GDA. Rg, response to gravity: light blue, most seedlings in category 1 or 2; intermediate blue, most in category 2 or 3; dark blue, most in category 3 or 4; deep blue, most in category 4; white, more seedlings scored in category 1 with GDA than without GDA (see Methods). Missing boxes, not tested.

Three conclusions follow: Hsp90 has an important role in many aspects of developmental plasticity; these roles vary in different genetic backgrounds; and the dependencies of different pathways on Hsp90 segregate independently of each other.

Hsp90 buffers developmental stability

Given the near homozygosity of inbred lines and accessions, the

partial penetrance of certain phenotypes was surprising. Although some morphogenetic traits were highly penetrant (for example, S-shaped rosettes in line 082 at 27°C), most were not. Similarly, responses of individual seedlings for some developmentally plastic traits (for example, germination and gravitropism) varied within a line. One explanation for this phenotypic variability is the segregation of genetic variation. Low levels of heterozygosity are present in RI lines (~1 site per 5,000 base pairs) and might also be present in accessions. Indeed, in one case tested, segregating genetic variation did affect penetrance. About 30% of Shadara seedlings exhibited deformed, missing or radially symmetrical true leaves with GDA. Without the drug, subtle abnormalities of true leaf development appeared more frequently in Shadara (~3–6%) than in other accessions (<1%). When plants displaying subtle abnormalities without GDA were self-bred, 50% of progeny exhibited the phenotype without GDA, and 75% did with GDA. Thus, genetic variants contributing to abnormal leaf development were still segregating in the Shadara seeds we sampled and could be enriched by selective breeding, increasing phenotypic penetrance. In contrast, in an RI line displaying partial penetrance of one altered phenotype (line CS1941, radially symmetrical and missing true leaves), self- and cross-breeding the affected progeny did not increase penetrance in the next generation (data not shown). Here, partial penetrance was not due to segregating genetic variation, suggesting that stochastic processes influenced the propensity of genomes to produce particular traits.

To investigate the interplay of genetic variation, stochastic processes, and their dependence on the Hsp90 buffering system, we tested crosses of different, nearly homozygous accessions (Col × Ler and Ler × Cvi) that are believed to have been recently evolving independently. If stochastic events contribute to phenotypes predisposed by genotype, then mixing genomes might disrupt developmental stability in F₁ progeny. Further, if Hsp90 buffers developmental stability against such events, reducing Hsp90 function in nearly identical F₁ progeny should increase phenotypic variance.

Indeed, F₁ progeny exhibited higher frequencies of subtle altered phenotypes than their parental accessions even without GDA (frequency 5–7%). In the presence of GDA, not only did the frequency of abnormal phenotypes increase dramatically (up to 25%), but the severity and complexity of the phenotypes also

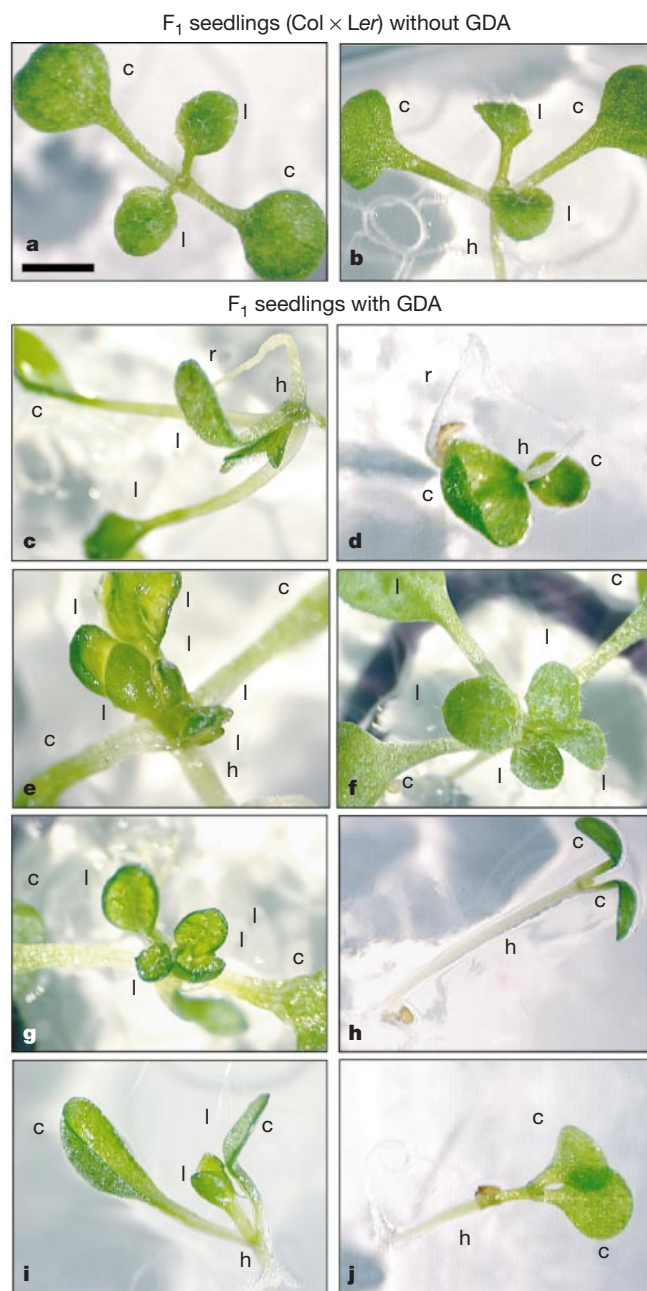


Figure 4 F₁ progeny of accession crosses exhibit higher diversity and greater phenotype complexity than parental accessions, especially with GDA. **a, b**, F₁ seedlings of Col × Ler cross without GDA. **c–j**, Seedlings with GDA. **c, d**, seedlings with hypocotyl curls and different cotyledon shapes. **e–g**, seedlings with multiple true leaves emerging apparently simultaneously: **e**, at different angles; **f**, with a leaf in first whorl missing; **g**, as seemingly separated rosettes. **d, h–j**, seedlings with diverse hypocotyl and cotyledon phenotypes: **d**, short curled hypocotyl, inverted orientation, epinastic cotyledons; **h**, long hypocotyl, small, dark-green cotyledons; **i**, large spoon-shaped cotyledons, extremely short hypocotyls; **j**, hypocotyl of intermediate size, light, non-expanded round cotyledons. Plant organs indicated, as defined in Fig. 1. Scale bar: 2 mm for **a–d, h–j**; 1 mm for **e–g**.

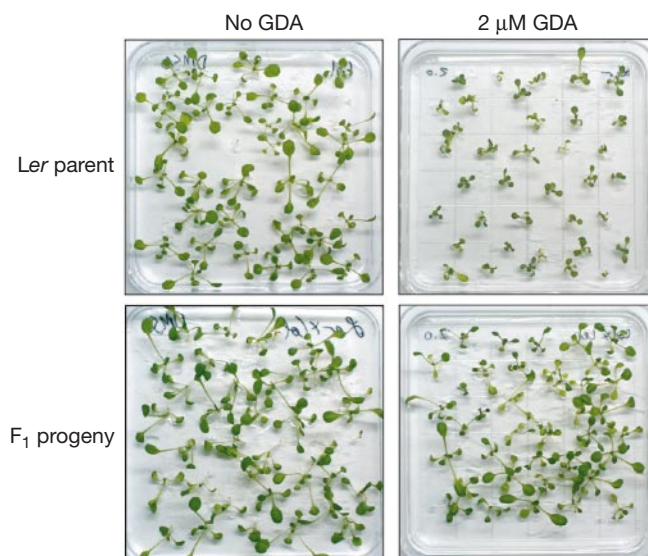


Figure 5 F₁ progeny grow more robustly in the presence of 2 μM GDA than do parental accessions. Left, without GDA; right, with GDA. Top, parental accession Ler; bottom, F₁ progeny of Col × Ler cross. Similar numbers of seeds (~40) were plated for all.

increased (Fig. 4). Phenotypes were stronger and more frequent in seedlings with a maternal contribution from *Ler* (~20–25% affected seedlings) than with maternal contributions from *Col* or *Cvi* (~10–15%).

Notably, the developmental diversity of F_1 seedlings with GDA was not due to reduced viability. When the concentration of GDA was doubled (to 2 μ M), *Col* $\times$ *Ler* F_1 seedlings exhibited an even greater abundance of phenotypes, but most were more vigorous and developmentally advanced than those of either parental accession at the same GDA concentration (Fig. 5). Significantly, when buffering functions were challenged by a higher growth temperature (27 °C) instead of by GDA, phenotypes of *Col* $\times$ *Ler* F_1 seedlings were also more diverse than those of their parental accessions. Again, the F_1 seedlings were generally larger and healthier (data not shown). Preliminary analysis of other cross-bred seedlings suggests that the effects of heterozygosity on developmental stability and stress tolerance vary, depending on the genetic divergence of the accessions and the nature of the stress, offering a rich subject for further analysis.

Discussion

Our work establishes three aspects of Hsp90 function in organismal biology. First, its role in buffering the expression of genetic variation⁶ is conserved across plant and animal kingdoms. Although it is difficult to predict what might prove adaptive in evolution, it is noteworthy that the phenotypes revealed in plants by challenging Hsp90 buffering capacity are not ‘monstrous’ in character (as they might be described in fruitflies). Indeed, some would seem plausibly advantageous under particular conditions: for example, altered leaf shapes, purple pigment accumulation, and different degrees of hypocotyl extension. Notably, most Hsp90-buffered traits we tested were also uncovered by moderate changes in growth conditions (that is, a mild temperature increase). This establishes that the Hsp90 buffer has global effects on the storage of cryptic polymorphisms and their release in response to shifting environments, converting neutral to non-neutral variation.

Second, Hsp90 profoundly affects developmental plasticity. Plasticity allows morphogenetic variation in response to circumstances and environments and is particularly salient in plants. Our observations suggest that Hsp90 functions at the interface between genotypes and environments, a unique vantage point from which to influence the dynamic nature of developmental processes. Most remarkably, the degree to which the plasticity of individual traits depends on Hsp90 varied greatly between RI lines, and different traits exhibited different levels of Hsp90 dependence within each RI line. Apparently, the genetic networks that affect both the likelihood that a particular trait might appear, as well as the extent to which it may respond to changes in the environment, are commonly polymorphic with respect to the influence of Hsp90. We suggest that, in the face of changing selective pressures, Hsp90 buffering may provide an avenue by which populations can evolve different genotypic states—from those that produce a particular trait, to those in which the trait dynamically responds to the environment, to those in which a different developmental endpoint has become a fixed characteristic.

Third, Hsp90 buffers developmental stability against stochastic processes. Individual seedlings in nearly isogenic RI lines and accessions produced strong phenotypes when Hsp90 function was reduced, but these were only partially penetrant. Also, the nearly genetically identical F_1 progeny of crosses between *Col* and *Ler* parents produced an unprecedented diversity of phenotypes when Hsp90 function was challenged. These phenotypes were not due to a general loss of vigour, because F_1 progeny from the same cross were generally more robust than either parent under conditions of increased stress. Thus, Hsp90 normally acts to reduce the likelihood that stochastic events will alter the deterministic unfolding of a multitude of developmental programmes.

These seemingly diverse effects of Hsp90 are readily encompassed by a single, simple molecular framework. The biochemical function of Hsp90 is to chaperone metastable proteins, including diverse regulators of growth and development. In doing so, Hsp90 stabilizes its ‘client’ proteins in conformations that would otherwise be prone to misfolding and potentiates their capacity to be activated in the proper time and place, by associations with partner proteins, ligand binding, post-translational modifications and correct localization^{7–9}. Under stressful conditions, Hsp90 is induced in response to problems in protein folding. Polymorphisms with the potential to alter a particular trait may be so enriched in individual genomes that Hsp90’s ability to maintain the functional pathway is exceeded. These polymorphisms might occur not only in client proteins themselves but also in upstream or downstream components of the pathway in which they function. Indeed, even in the absence of stress, some populations may be so close to the trait expression threshold that stochastic events in development will produce a few individuals displaying the altered trait. Once the pathway is diverted, the expression of a new trait may become robust through the influence of auto-regulatory feedback loops, self-perpetuating protein conformations, and developmental windows. Thus, Hsp90 contributes to phenotypic variance by buffering the functional state (folding, accumulation and local concentration) of gene products that contribute to altered traits through the effects of chance, genotypes and environments.

From a the viewpoint of a population geneticist, buffering systems like Hsp90, without regard to any possible adaptive value, would decrease selection on nucleotide substitutions, allowing storage of an expanded spectrum of selectively nearly neutral ones. Under stable environmental conditions, a population arrives at a local fitness optimum in an adaptive landscape. Given that natural selection can only further increase the fitness of a population, it is a perplexing evolutionary question how a population might move to a different local optimum without an intervening period of reduced fitness (adaptive valley). Previously proposed mechanisms include genetic drift in small populations, compensatory mutations and gene conversion^{25–27}. As a by-product of its biochemical function, Hsp90 may allow the neutral accumulation of potentially selectable polymorphisms and synchronize their conversion to a non-neutral state. Certainly, most combinations of polymorphisms will be deleterious, and these may be periodically purged from the population through the environmental coupling of the Hsp90 buffer. However, rare combinations may produce a new, advantageous phenotype, thereby providing a molecular means by which adaptive peak shifts in large populations may occur without passing through an adaptive valley. None of the other mechanisms discussed can be modulated by environmental contingencies.

Finally, deliberate manipulation of the Hsp90 buffer may offer an opportunity to speed the identification of genetic variation for academic investigations and commercial applications. Developmental pathways might be deciphered more rapidly by using multiple, subtle polymorphisms contributing to a trait than by traditional approaches of sequentially identifying strong alleles, enhancers and suppressors. Harnessing such variation to produce valuable phenotypes would also bypass public concern about transgenic plants. Our results also suggest possible connections to medical maladies with stochastic genotypic and environmental aetiology, because they provide a global molecular explanation for why genotype does not always directly translate into phenotype. Indeed, our findings situate Hsp90 at the critical intersection of genotype, environment and development. □

Methods

Plant material and growth conditions

Accession seeds (Mr-0, Shadara, Ts-1, Tsu-1) and RI lines (CS22477, CS1899) were from the ABRC. Columbia (*Col*-0), *Ler* and *Cvi* seeds were laboratory stocks propagated from seeds (D. Preuss).

Accessions were independently cross-fertilized several times to produce F₁ seeds. To minimize variation in growth conditions, surface-sterilized seeds were plated with equal spacing on germination (GM) medium²² containing either 1 μ M GDA (C₃₀H₄₂N₂O₈, formula weight 559, from 1 mg ml⁻¹ stock in dimethyl sulphoxide, DMSO) or equivalent concentration of DMSO alone (0.056%). Plates were wrapped in foil to allow germination while avoiding degradation of light-sensitive GDA, and incubated at 22 °C after cold treatment for 24–48 h. Plates were unwrapped after 48 h at 22 °C, incubated in continuous light at ~150 μ mol m⁻² s⁻¹, and analysed after 8 and 10 days. Digital images were taken after 10–12 days (Figs 1, 2 and 4) and after 14 days (Fig. 5) (at $\times 10$ or $\times 20$ magnification, as indicated).

Experiments in altered environments

Incubation at 27 °C increased growth rates (by ~0.5–1 day after 8 days), and GDA decreased growth rates (by ~0.5–1 day) relative to plants grown without GDA at 22 °C. To accommodate these differences, one set of seeds (~28 per RI line) was plated with GDA on day 1; a second set (22 °C) was plated without GDA on day 2. On day 3, three sets were plated: one for growth at 27 °C and two at 22 °C, with and without GDA. To minimize variation and provide more objective comparisons, two RI lines were plated together on each plate. Sets were cold treated for 48 h. Foil-wrapped plates incubated at 22 °C were treated as described above. Because development was more rapid at 27 °C, these plates were unwrapped after 24 h. Phenotypes were scored at 8 and 10 days.

Developmental plasticity responses

Seedlings respond to subtle environmental changes that are difficult to control (for example, small variations in incubator conditions, or slight substratum changes). Values for individual lines varied somewhat between experiments, but within experiments seedlings on multiple replicate plates behaved similarly. For example, for the dark response of hypocotyls and roots, 50 seeds for each RI line were plated on vertical plates (10 seeds per plate) and intermixed with multiple RI lines in the incubator. A second screen for the dark response analysed ten seeds for each RI line with a 2-h light pulse and 48 h of cold treatment. Hypocotyl length reductions were similar to those of Fig. 3 for nearly all RI lines (hypocotyls of lines 117 and 086 were much longer than those of light-grown seedlings).

After 7 days in the dark, hypocotyl and root lengths were measured on digitized images using Scion Image (<http://www.scioncorp.com>). The number of germinated seeds and morphological characteristics of seedlings grown with and without GDA were documented. Germination efficiency of Cvi seeds increased (up to 100%) with a 2-h light pulse. This light pulse also increased germination efficiencies of lines indicated with white or yellow boxes in Fig. 3, except line 134.

The light-induced rate of GDA decay was assessed phenotypically. GDA-containing plates were pre-exposed to light for various periods before seeds were plated, and hypocotyl growth was measured after 7 days in the dark, as in Fig. 3. Exposures of 2–4 h had little effect on GDA potency; 48 h eliminated it.

Because GDA affects root extension, the response to changing the direction of gravity was tested at GDA concentrations of 0.75 and 1 μ M, ensuring that a set of seedlings was available for each RI line with root growth comparable to controls. Forty seeds for each condition were plated, cold treated for 48 h, and incubated at 22 °C in the dark for 4 days (to exclude phototropism signals). Root growth and germination were scored, and plates were turned by 90°. Responses were scored 12, 24, 36 and 48 h after turning. Each seedling was assigned a turning category (1, complete turn; 2, incomplete turn; 3, no turn; 4, turning upward; 5, coiled growth). Line responses (blue shading, Fig. 3) were scored according to the number of seedlings in each category. Seedlings of most RI lines without GDA scored in category 1, except lines 129 and 131, which were distributed between categories 1 and 2.

Received 13 December 2001; accepted 15 April 2002; doi: 10.1038/nature749.

Published online 12 May 2002.

1. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).
2. Mather, K. Genetical control of stability in development. *Heredity* **7**, 297–336 (1953).
3. Durrant, A. The environmental induction of heritable change in *Linum*. *Heredity* **17**, 27–61 (1962).
4. McLaren, A. Too late for the midwife toad: stress, variability and Hsp90. *Trends Genet.* **15**, 169–171 (1999).
5. Waddington, C. H. Genetic assimilation of an acquired character. *Evolution* **7**, 118–126 (1952).

6. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
7. Buchner, J. Hsp90 & Co.—a holding for folding. *Trends Biochem. Sci.* **24**, 136–141 (1999).
8. Mayer, M. P. & Bukau, B. Molecular chaperones: the busy life of Hsp90. *Curr. Biol.* **9**, R322–R325 (1999).
9. Young, J. C., Moarefi, I. & Hartl, F. U. Hsp90: a specialized but essential protein-folding tool. *J. Cell Biol.* **154**, 267–274 (2001).
10. Abbott, R. J. & Gomes, M. F. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62**, 411–418 (1989).
11. Kuittinen, H., Mattila, A. & Savolainen, O. Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. *Heredity* **79**, 144–152 (1997).
12. Callahan, H. S., Pigliucci, M. & Schlichting, C. D. Developmental phenotypic plasticity: where ecology and evolution meet molecular biology. *Bioessays* **19**, 519–525 (1997).
13. Pigliucci, M. & Byrd, N. Genetics and evolution of phenotypic plasticity to nutrient stress in *Arabidopsis*: drift, constraints or selection? *Biol. J. Linn. Soc.* **64**, 17–40 (1998).
14. Stratton, D. A. Reaction norm functions and QTL–environment interactions for flowering time in *Arabidopsis thaliana*. *Heredity* **81**, 144–155 (1998).
15. Dorn, L. A., Pyle, E. H. & Schmitt, J. Plasticity to light cues and resources in *Arabidopsis thaliana*: testing for adaptive value and costs. *Evolution* **54**, 1982–1994 (2000).
16. Milioni, D. & Hatzopoulos, P. Genomic organization of hsp90 gene family in *Arabidopsis*. *Plant Mol. Biol.* **35**, 955–961 (1997).
17. Krishna, P. & Gloor, G. The Hsp90 family of proteins in *Arabidopsis thaliana*. *Cell Stress Chaperon.* **6**, 238–246 (2001).
18. Roe, S. M. *et al.* Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. *J. Med. Chem.* **42**, 260–266 (1999).
19. Dutta, R. & Inouye, M. GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem. Sci.* **25**, 24–28 (2000).
20. Alonso-Blanco, C. *et al.* Development of an AFLP based linkage map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. *Plant J.* **14**, 259–271 (1998).
21. Mayfield, J. A., Fiebig, A., Johnstone, S. E. & Preuss, D. Gene families from the *Arabidopsis thaliana* pollen coat proteome. *Science* **292**, 2482–2485 (2001).
22. Queitsch, C., Hong, S. W., Vierling, E. & Lindquist, S. Heat shock protein 101 plays a crucial role in thermotolerance in *Arabidopsis*. *Plant Cell* **12**, 479–492 (2000).
23. Donohue, K., Messiqua, D., Pyle, E. H., Heschel, M. S. & Schmitt, J. Evidence of adaptive divergence in plasticity: density- and site-dependent selection on shade-avoidance responses in *Impatiens capensis*. *Evolution* **54**, 1956–1968 (2000).
24. Ballare, C. L. Keeping up with the neighbours: phytochrome sensing and other signalling mechanisms. *Trends Plant Sci.* **4**, 97–102 (1999).
25. Kimura, M. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc. Natl Acad. Sci. USA* **88**, 5969–5973 (1991).
26. Ludwig, M. Z., Bergman, C., Patel, N. H. & Kreitman, M. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564–567 (2000).
27. Hansen, T. F., Carter, A. J. & Chiu, C. H. Gene conversion may aid adaptive peak shifts. *J. Theor. Biol.* **207**, 495–511 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank H. Oakley, A. Evans and the University of Chicago greenhouse staff for technical help. We are especially grateful to J. Malamy, J. Greenberg, D. Preuss, J. Borevitz, J. Maloof and members of the Lindquist laboratory for advice and discussions, and to N. Patel for use of equipment and comments. We thank the ABRC for seeds. T.A.S. is a Howard Hughes Medical Institute pre-doctoral fellow. S.L. thanks the Howard Hughes Medical Institute and the National Institutes of Health for their support.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.L. (e-mail: lindquist_admin@wi.mit.edu).

Observational evidence for the accretion-disk origin for a radio jet in an active galaxy

Alan P. Marscher*, Svetlana G. Jorstad*†, José-Luis Gómez‡, Margo F. Aller§, Harri Teräsanta||, Matthew L. Lister¶ & Alastair M. Stirling#

* Institute for Astrophysical Research, Boston University, 725 Commonwealth Avenue, Boston, Massachusetts 02215, USA

† Sobolev Astronomical Institute, St Petersburg State University, Universitetskii Prospekt 28, Petrodvorets, St Petersburg, 198504, Russia

‡ Instituto de Astrofísica de Andalucía (CSIC), Apartado Correos 3004, Granada, E-18080, Spain

§ Astronomy Department, University of Michigan, 830 Dennison, 501 East University Street, Ann Arbor, Michigan 48109-1090, USA

|| Metsähovi Radio Observatory, Helsinki University of Technology, Metsähovintie 114, Kylmälä, Finland-02540

¶ National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, Virginia 22903-2475, USA

Center for Astrophysics, University of Central Lancashire, Preston, Lancashire, PR1 2HE, UK

Accretion of gas onto black holes is thought to power the relativistic jets of material ejected from active galactic nuclei (AGN) and the ‘microquasars’ located in our Galaxy^{1–3}. In microquasars, superluminal radio-emitting features appear and propagate along the jet shortly after sudden decreases in the X-ray fluxes¹. This establishes a direct observational link between the black hole and the jet: the X-ray dip is probably caused by the disappearance of a section of the inner accretion disk⁴ as it falls past the event horizon, while the remainder of the disk section is ejected into the jet, creating the appearance of a superluminal bright spot⁵. No such connection has hitherto been established for AGN, because of insufficient multi-frequency data. Here we report the results of three years of monitoring the X-ray and radio emission of the galaxy 3C120. As has been observed for microquasars, we find that dips in the X-ray emission are followed by ejections of bright superluminal knots in the radio jet. The mean time between X-ray dips appears to scale roughly with the mass of the black hole, although there are at present only a few data points.

We selected 3C120 (redshift $z = 0.033$) owing to the combination of superluminal motion in its radio jet^{6,7} and its Seyfert-like X-ray properties. The X-ray energy spectrum^{8–11} (flux density as a function of frequency) is well fitted by a power law of slope $\alpha_E = -0.7$ to -1 , plus an iron emission line at a photon energy of 6.4 keV. The prevailing models¹¹ explain the X-ray emission from Seyferts in a similar manner as that from microquasars⁴: soft X-rays from the inner accretion disk plus harder X-rays from a corona of hot electrons that reprocess the disk radiation. The presence of the iron line implies that most of the X-ray flux arises from the region near the accretion disk rather than in the jet. The similarity between its X-ray emission and that of microquasars, plus the presence of superluminal motion in the radio jets of both, led us to search for a connection between X-ray events and superluminal ejections in 3C120.

We imaged the compact radio jet of 3C120 with the Very Long Baseline Array (VLBA) roughly monthly from November 1997 to March 1999 (refs 6, 7), and bimonthly thereafter. A description of the data analysis procedures is provided in refs 6 and 7. The sequence of images between March 1998 and April 2001 is shown in Fig. 1. Diagonal lines indicate the trajectories of all major superluminal knots that emerged from the core during the period of our X-ray monitoring programme. Each of these is distinguished

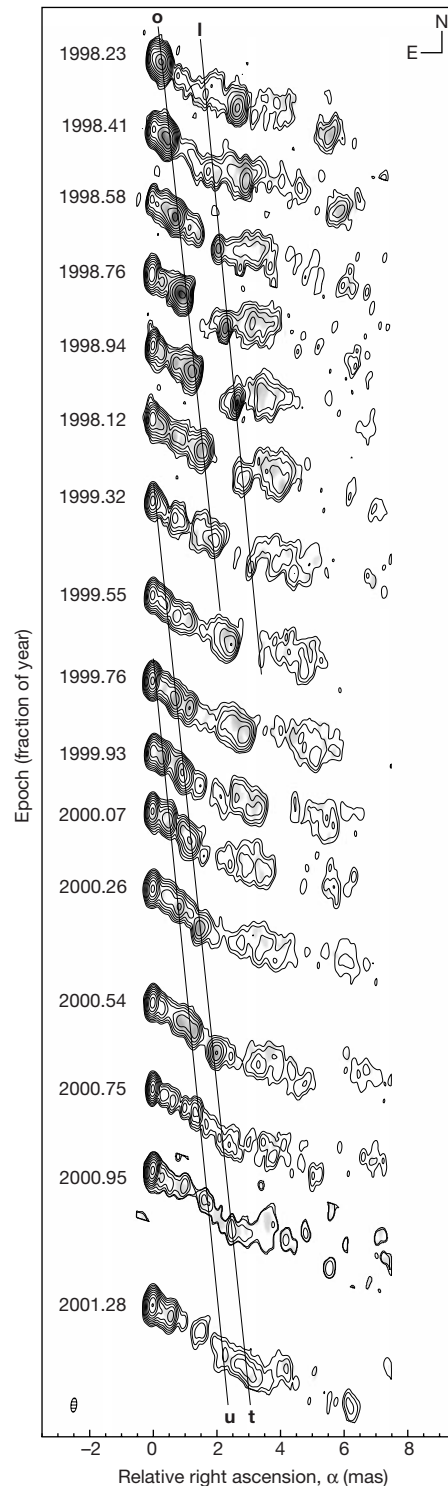


Figure 1 Sequence of VLBA images of 3C120 at a frequency of 43 GHz. Time progresses linearly in the downward direction; for each image, north is upward and east to the left. Contours represent total intensity, with values of 0.5, 1, 2, 4, 8, 16, 32, 64 and 90% of the peak of 1.52 Jy per beam (which occurred at epoch 1998.23). Contours increase in intensity towards the centre. Greyscale corresponds to polarized intensity, with a maximum of 36 mJy per beam (at 1998.94). The small ellipse in the bottom left corner is the restoring beam, which approximates the angular resolution in different directions. Diagonal lines trace the approximate path of the superluminal knots, which are identified by the letters l, o, t and u, following the sequence of refs 6 and 7. At the distance of 3C120 of 140 Mpc (for a Hubble constant with value $65 \text{ km s}^{-1} \text{ Mpc}^{-1}$), 1 mas corresponds to a length of 0.70 pc projected on the plane of the sky.

from other features in the jet by high flux density and significant linear polarization. Their apparent velocities range from 4.1 to $5.0c$ (where c is the speed of light, and for a Hubble constant of $65 \text{ km s}^{-1} \text{ Mpc}^{-1}$). We trace the motions to extrapolated 'ejection' times when each knot was coincident with the 'core', the bright spot on the eastern end of the jet.

We observed 3C120 in the 2.4–20-keV X-ray band from the end of 1997 to early 2000 with the *Rossini* X-ray Timing Explorer (RXTE). The observations were carried out weekly in 1998 and semi-weekly in 1999, except for 8-week periods when the position of 3C120 was too close to the Sun. To complete the X-ray light curve, shown in Fig. 2a, we processed archival data to add daily RXTE measurements in early 1997. To show the connection between the radio and X-ray emission, we display in Fig. 2 the dates of superluminal ejections as well as the variation with time of the X-ray spectral index α_E , total flux density at three radio frequencies, and flux density of the core at 43 GHz (from the images). Each superluminal ejection (with the possible exception of event t) coincides with the onset of an increase in flux density in at least one of the radio light curves.

We designed our RXTE programme to search for X-ray dips. In order to define such events, we first divide the data into two segments: (A), early 1997; and (B), the remainder. (The Mann–Whitney test¹² indicates that the X-ray fluxes from the two sections

are not drawn from the same distribution, with 1.5% significance.) We identify dips as intervals when ≥ 4 consecutive fluxes lie more than 1σ from the mean of the relevant segment of data. We thus identify four dips in the X-ray light curve, denoted in red in Fig. 2a, b.

The Shapiro–Wilk¹³ test shows that the distribution of fluxes from segment B is consistent with a gaussian distribution. However, this is not the case for segment A unless the outliers (the first dip and the apparent flare near the start of 1997) are removed. In order to test the statistical significance of an observed event consisting of n flux measurements in a data set containing N values, we calculate the probability $P(n, N, F_d)$ that n consecutive fluxes would fall below a level F_d by random chance, where F_d is the highest point in each dip. We determine the probability $p(F \leq F_d)$ from the best-fit gaussian curve for segment B and by rank-order (the i th lowest flux out of 52 data points has $p = i/52$) for segment A. A recursion formula¹⁴ gives $P(n, N, F_d) = 0.011, 1.2 \times 10^{-3}, 0.010$ and 1.8×10^{-5} for the four events. Each event is therefore only very rarely expected to occur by random chance. The significance of each dip is further supported by flattening of the X-ray spectrum (see Fig. 2b), as occurs also in the case of the microquasar GRS 1915+105 (ref. 15).

It is apparent from Fig. 2 that each X-ray dip is followed shortly by the appearance of a superluminal knot at the site of the radio core. The latter occurred at the following epochs, obtained by a straight line fit to the separation-from-core versus time dependence for each knot: 1997.25 ± 0.04 , 1998.08 ± 0.03 , 1999.27 ± 0.03 , and 1999.70 ± 0.02 . To estimate the probability that the X-ray dips and superluminal ejections are associated by random chance, we consider that there are about two years of X-ray data and a ~ 0.2 -yr window within which we would consider the events to be related. The probability of any single radio event occurring randomly less than 0.2 yr after a particular X-ray event is thus about 0.1. There are four events of each type, so the probability of all four radio and X-ray events being close in time by chance is thus about $4!(0.1)^4 \approx 0.2\%$. We therefore reject the null hypothesis.

The mean time delay between the minimum in the X-ray flux and the time of ejection is 0.10 ± 0.03 yr. At a typical superluminal apparent speed of $4.7c$, a knot moves a distance of 0.14 parsecs (pc) in 0.10 yr, projected on the plane of the sky. If we adopt an angle of the jet to the line of sight of 20° (the maximum allowed given the apparent velocity⁶; this minimizes the de-projected distance), the actual distance travelled is 0.41 pc. However, it may be the case that the jet starts as a broad wind that accelerates and becomes increasingly focused until it reaches the radio core^{16–18}, in which case we calculate that a correction factor of around 0.8 must be applied. We therefore derive a minimum distance of 0.3 pc from the X-ray source to the core of the radio jet. This corresponds to a projected angular displacement of 0.16 milliarcseconds (mas) eastward of the core in the images of Fig. 1. We note that the core of the radio jet is not coincident with the black hole, as has been assumed¹⁸ in the case of the radio galaxy M87. Rather, it might be where the jet flow reaches its asymptotic relativistic velocity¹⁶ or the site of standing shocks that accelerate electrons to ultrahigh energies^{5,19}. This result implies that the jet might be essentially invisible between its site of origin near the black hole³ and the radio core far downstream where its plasma is re-energized.

The correlated X-ray and radio behaviour of 3C120 is similar to that of the microquasar GRS1915+105: X-ray dips during which the spectrum hardens, each followed by the ejection of a bright superluminal radio knot. This phenomenon has been interpreted¹⁴ as resulting from an instability in the accreting flow causing a piece of the inner disk to break off. Part of this matter is drawn into the event horizon of the black hole but with considerable matter and energy ejected down the jet. The loss of some of the inner disk causes a drop in the soft X-ray flux, which is observed as a dip plus hardening of the spectrum. The hard X-ray flux, presumably the product of

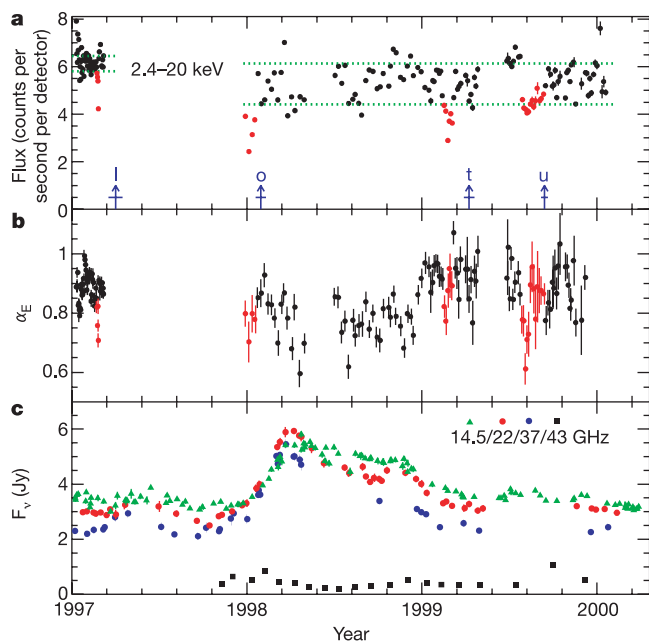


Figure 2 Time dependence of the X-ray and radio emission of 3C120. **a**, X-ray light curve in photon counts per second per detector; for many data points, the error bars are smaller than the symbols. For each observation, we determined the X-ray flux in photon counts per second over the energy range 2.4 to 20 keV by subtracting an X-ray background model (supplied by the RXTE Guest Observer Facility) from the raw spectrum, using the standard X-ray data analysis software package FTOOLS. We fitted the photon spectrum (observed flux F_{obs} versus energy E) with a model consisting of a power-law intrinsic spectrum, $F \propto E^{-\alpha}$, plus photoelectric absorption along the line of sight (the latter is not very important at these energies) using the program XSPEC. Red data points correspond to events that we identify as dips (in **a** and **b**), and the green dotted lines delineate $\pm 1\sigma$ of the mean of each segment of data. Times of ejections of superluminal radio knots are indicated by upward blue arrows, with the horizontal crossbar giving the uncertainty and the labels l, o, t and u corresponding to the features marked in Fig. 1. **b**, X-ray 'energy' spectral index α_E , where $-\alpha_E = \alpha - 1$ is the exponent of the power-law dependence of flux density on frequency. **c**, Radio light curves at 14.5 GHz (green triangles) (University of Michigan Radio Astronomy Observatory); 22 GHz (red circles) and 37 GHz (blue circles) (Metsähovi Radio Observatory); and 43 GHz (core only, black squares) (VLBA). Ref. 6 and this work.

Compton scattering—for example, in a corona above the disk^{1,4,11}—is not affected as much. According to this picture, the time delay between the X-ray dip and the appearance of the superluminal knot corresponds to the travel time of the disturbance from the inner accretion disk to the core of the radio jet.

Another property shared by 3C120 and GRS1915+105 is the bulk velocity of the jet needed to explain the apparent superluminal motion: $0.98c$ in each case (Lorentz factor ≈ 5)^{6,20}. The one-sided observed jet in 3C120 contrasts with the two oppositely directed jets in the microquasar. This can readily be explained as a consequence of the stronger relativistic beaming of the approaching jet in 3C120, because the angle between the jet and the line of sight is smaller ($<20^\circ$ compared with 66° in GRS1915+105)^{6,20}.

The radio galaxy 3C120 is therefore similar to a scaled-up version of a microquasar. It may be possible to adapt models for the closer microquasars—whose accretion processes are better studied owing to higher and more rapidly variable X-ray fluxes—to their distant AGN cousins. We expect the scaling factor to be proportional to the radius of the innermost orbit around the black hole³, which ranges from $3R_S$ for a static black hole to $0.5R_S$ for the maximally rotating case²¹. Here, $R_S \propto M$ is the Schwarzschild radius, with the mass of the black hole $M = 1.6\text{--}5.1 \times 10^7$ solar masses in 3C120 (ref. 22) and $10\text{--}14$ solar masses in GRS1915+105 (ref. 2). The distance of the core of the radio jet to the black hole is about 7×10^4 times higher in 3C120 than in GRS1915+105 (ref. 20), whereas the time interval between X-ray dips is about 5×10^5 times longer ($0.5\text{--}1.2$ yr for 3C120 versus $25\text{--}100$ s for the microquasar⁴). These values are, respectively, $20\text{--}70$ and $2\text{--}10$ times lower than the ratio of black hole masses. The factor of $2\text{--}10$ is small compared to the uncertainties in our theoretical understanding of how the time between events relates to the physical parameters of the accretion disk. The greater discrepancy in the ratio of length scales implies that the conditions governing formation of the radio core—perhaps the pressure of the external medium or the magnetic field in the jet—are different in the two objects. □

Received 15 January; accepted 26 April 2002; doi:10.1038/nature00772.

- Mirabel, I. F. & Rodríguez, L. F. Microquasars in our Galaxy. *Nature* **392**, 673–676 (1998).
- Greiner, J., Cuby, J. G. & McCaughrean, M. J. An unusually massive stellar black hole in the Galaxy. *Nature* **414**, 522–525 (2001).
- Meier, D. L., Koide, S. & Uchida, Y. Magnetohydrodynamic production of relativistic jets. *Science* **291**, 84–92 (2000).
- Belloni, T. Inner disk oscillations. *Astrophys. Space Sci.* **276** (suppl.), 145–152 (2001).
- Gómez, J. L., Martí, J. M., Marscher, A. P., Ibáñez, J. M. & Alberdi, A. Hydrodynamical models of superluminal sources. *Astrophys. J.* **482**, L33–L36 (1997).
- Gómez, J. L., Marscher, A. P., Alberdi, A., Jorstad, S. G. & García-Miró, C. Flashing superluminal components in the jet of the radio galaxy 3C120. *Science* **289**, 2317–2320 (2000).
- Gómez, J. L., Marscher, A. P., Alberdi, A., Jorstad, S. G. & Agudo, I. Monthly 43 GHz VLBA polarimetric monitoring of 3C 120 over 16 epochs: evidence for trailing shocks in a relativistic jet. *Astrophys. J.* **561**, L161–L164 (2001).
- Grandi, P. et al. ROSAT, ASCA, and OSSE observations of the broad-line radio galaxy 3C 120. *Astrophys. J.* **487**, 636–643 (1997).
- Halpern, J. P. X-ray spectrum and variability of 3C 120. *Astrophys. J.* **290**, 130–135 (1985).
- Maraschi, L. et al. Coordinated X-ray, ultraviolet, and optical observations of 3C 120. *Astrophys. J.* **368**, 138–151 (1991).
- Zdziarski, A. & Grandi, P. The broadband spectrum of 3C 120 observed by BeppoSAX. *Astrophys. J.* **551**, 186–196 (2001).
- Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947).
- Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611 (1965).
- Feller, W. *An Introduction to Probability Theory and its Applications*, 2nd edn 322 (Wiley, New York, 1968).
- Klein-Wolt, M. et al. Hard X-ray states and radio emission in GRS 1915+105. *Mon. Not. R. Astron. Soc.* **331**, 745–764 (2002).
- Marscher, A. P. Relativistic jets and the continuum emission in QSOs. *Astrophys. J.* **235**, 386–391 (1980).
- Reynolds, S. P. Theoretical studies of compact radio sources. I. Synchrotron radiation from relativistic flows. *Astrophys. J.* **256**, 13–27 (1982).
- Junor, W., Biretta, J. A. & Livio, M. Formation of the radio jet in M87 at 100 Schwarzschild radii from the central black hole. *Nature* **401**, 891–892 (1999).
- Daly, R. D. & Marscher, A. P. The gas dynamics of compact relativistic jets. *Astrophys. J.* **334**, 539–551 (1988).
- Fender, R. P. et al. MERLIN observations of relativistic ejections from GRS 1915+105. *Mon. Not. R. Astron. Soc.* **304**, 865–876 (1999).

- Shapiro, S. I. & Teukolsky, S. A. *Black Holes, White Dwarfs and Neutron Stars: the Physics of Compact Objects* 362 (Wiley Interscience, New York, 1983).
- Wandel, A., Peterson, B. M. & Malkan, M. A. Central masses and broad-line region sizes of active galactic nuclei. I. Comparing the photoionization and reverberation techniques. *Astrophys. J.* **526**, 579–591 (1999).

Acknowledgements

We gratefully acknowledge partial support for this research from the US National Science Foundation (NSF), NASA, the Fulbright Commission for collaboration between Spain and the USA, and the Spanish Dirección General de Investigación Científica Técnica. The VLBA is an instrument of the National Radio Astronomy Observatory, a facility of the NSF operated under cooperative agreement by Associated Universities Inc. The University of Michigan Radio Astronomy Observatory was funded in part by the NSF and by the University of Michigan Department of Astronomy.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.P.M. (e-mail: marscher@bu.edu).

Coherence–incoherence and dimensional crossover in layered strongly correlated metals

T. Valla*, P. D. Johnson*, Z. Yusuf†, B. Wells†, Q. Li‡, S. M. Loureiro§, R. J. Cava§, M. Mikami||, Y. Mori||, M. Yoshimura|| & T. Sasaki||

* Physics Department, † Material Sciences Department, Brookhaven National Laboratory, Upton, New York 11973, USA

‡ Department of Physics, University of Connecticut, 2152 Hillside Road U-46, Storrs, Connecticut 06269, USA

§ Department of Chemistry and Princeton Materials Institute, Princeton University, Princeton, New Jersey 08540, USA

|| Department of Electrical Engineering, Osaka University, 2-1 Yamada-oka, Suita-shi, Osaka, 565-0871 Japan

The properties of an interacting electron system depend on the electron correlations and the effective dimensionality. For example, Coulomb repulsion between electrons may inhibit, or completely block, conduction by intersite electron hopping, thereby determining whether a material is a metal or an insulator¹. Furthermore, correlation effects increase as the number of effective dimensions decreases; in three-dimensional systems, the low-energy electronic states behave as quasiparticles, whereas in one-dimensional systems, even weak interactions break the quasiparticles into collective excitations². Dimensionality is particularly important for exotic low-dimensional materials where one- or two-dimensional building blocks are loosely connected into a three-dimensional whole. Here we examine two such layered metallic systems with angle-resolved photoemission spectroscopy and electronic transport measurements, and we find a crossover in the number of effective dimensions—from two to three—with decreasing temperature. This is apparent from the observation that, in the direction perpendicular to the layers, the materials have an insulating character at high temperatures but become metal-like at low temperatures, whereas transport within the layers remains metallic over the whole temperature range. We propose that this change in effective dimensionality correlates with the presence of coherent quasiparticles within the layers.

Crossover in interlayer transport has been detected in layered metals such as Sr_2RuO_4 (ref. 3) and NaCo_2O_4 (ref. 4), and more recently in $(\text{Bi}_{1-x}\text{Pb}_x)_2\text{M}_3\text{Co}_2\text{O}_y$ (where M is Ba or Sr)^{5,6}. The layers appear ‘isolated’ at high temperatures, but become connected at low

temperatures to give a three-dimensional (3D) system. A similar crossover is observed in the quasi-one-dimensional (1D) Bechgaard salt (TMTSF)₂PF₆ in the least-conductive direction⁷. The crossover temperature, T_M , is typically between 90 and 200 K. The low-temperature phase (temperature $T \ll T_M$) is 3D-like in the sense that the resistivity has nearly the same temperature dependence in all three directions.

The transport in the c -axis direction (perpendicular to the layers) of an anisotropic layered system may be coherent or incoherent. In a system showing coherent transport, the electronic states are well described in terms of the dispersing 3D momentum states in the usual band formulation. The conductivity is determined by the scattering rate $\Gamma = 1/\tau$ (where τ is the quasiparticle lifetime) and the Fermi velocity in that direction. The system may be anisotropic, but with a nearly temperature-independent anisotropy ratio, that is, $\rho_c(T) \propto \rho_{ab}(T) \propto \Gamma(T)$, where ρ_{ab} and ρ_c are the in-plane and interlayer resistivities, respectively. In the case of incoherent transport, when the quasiparticle scattering rate is much greater than the effective interlayer hopping, the interlayer tunnelling events are uncorrelated. Formally, this means that the electrons are scattered many times between successive tunnelling events and the perpendicular momentum, k_z , is not a good quantum number. The conductivity, σ_c , is then proportional to the tunnelling rate between two adjacent layers, and assuming that the intralayer momentum is conserved in a tunnelling event^{8,9}:

$$\sigma_c(T, \omega = 0) \propto \int \frac{d^2k}{(2\pi)^2} t_{\perp}(k)^2 G_R(k, \omega) G_A(k, \omega) \frac{\partial f(\omega)}{\partial \omega} \quad (1)$$

where $G_{R,A}$ is the retarded or advanced in-plane Green's function near the Fermi level ($\omega = 0$), where ω is the binding energy, $t_{\perp}(k)$ is the interplane hopping and f is the Fermi distribution. In the quasiparticle picture, the Green's functions have a coherent component $\sim Z/(\omega - \varepsilon_k - i\Gamma)$, where Z is the quasiparticle renormalization factor and ε_k is the renormalized quasiparticle energy. This results in the interlayer transport (equation (1)) having the same temperature dependence as the in-plane transport, dictated by

$\Gamma(T)$. The crossover behaviour that we report here therefore suggests that the quasiparticle picture is inappropriate above T_M , where conductivities are uncoupled⁸.

Angle-resolved photoelectron spectroscopy (ARPES) is an ideal experimental probe for testing the character of excitations. It measures directly the same single-particle spectral function $A(k, \omega) \propto \text{Im } G(k, \omega)$ that enters the transport equation (1). However, in contrast to transport probes, ARPES is k -resolving and able to probe deeper states that may contribute to the transport only at higher temperatures, $k_B T \approx \omega$. Applying ARPES studies to these materials, we find that the crossover observed in transport strongly correlates with changes observed in the spectral function.

In Fig. 1 we show the photoemission intensity recorded from (Bi_{0.5}Pb_{0.5})₂Ba₃Co₂O₇ in the (0,0) to (π, π) direction of the Brillouin zone. This material is a non-superconducting derivative of the double-layer copper oxide superconductor Bi₂Sr₂CaCu₂O_{8+ δ} , where the Cu–O planes are substituted by Co–O. Resistivity measurements on samples from the same batch (Fig. 2b) give metallic ρ_{ab} , while ρ_c shows a crossover at $T_M \approx 200$ K. The anisotropy ρ_c/ρ_{ab} increases with decreasing temperature, saturating below ~ 150 K at a value of about 7×10^3 .

The wide-range, low-temperature spectrum (Fig. 1a) shows a broad (~ 0.8 eV full-width at half-maximum, FWHM) hump, centred at ~ 0.6 eV, and a sharp state close to the Fermi level with a dip in between. The sharp peak disperses with parallel momentum, $k_{\parallel}$, and crosses the Fermi level with a small velocity $V_F \approx 160$ meV Å⁻¹ (note that this is one of the smallest Fermi velocities measured in ARPES), as visible in the expanded low-energy region of the contour spectrum (Fig. 1b). Our detailed studies indicate that this observation does not depend significantly on the in-plane azimuth. The sharp state forms a large, nearly cylindrical (Fermi wavevector, $k_F \approx 0.5$ Å⁻¹), Fermi surface centred at the Γ -point, which remains ungapped down to the lowest measured temperature, in accord with the metallic character of Co–O planes. The state is well defined only at low temperatures and only within the range of ~ 40 – 50 meV from the Fermi level. At higher temperatures, when

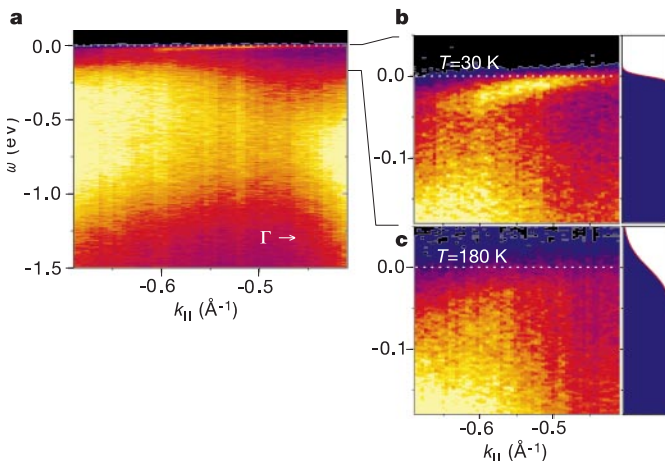


Figure 1 Angle-resolved photoelectron spectra of (Bi_{0.5}Pb_{0.5})₂Ba₃Co₂O₇. **a**, The contour plot of the ARPES intensity along the (0,0) to (π, π) line at $T = 30$ K. The arrow indicates the direction to the zone centre Γ . Changes in the low-energy region with temperature: the 30 K (**b**) and 180 K (**c**) contour spectra (left) are plotted on the same scale with the corresponding Fermi distributions (right). The spectra were taken with a Scienta SES200 hemispherical analyser with an angular resolution of $\pm 0.1^\circ$ or better and an energy resolution of the order of 10 meV. Photons of 15.4 eV, provided by a normal incidence monochromator based at the NSLS, were used for the excitation. Samples, grown as described in ref. 5, were mounted on a liquid He cryostat, and cleaved *in situ* in the ultrahigh-vacuum chamber with base pressure $(2\text{--}3) \times 10^{-9}$ Pa.

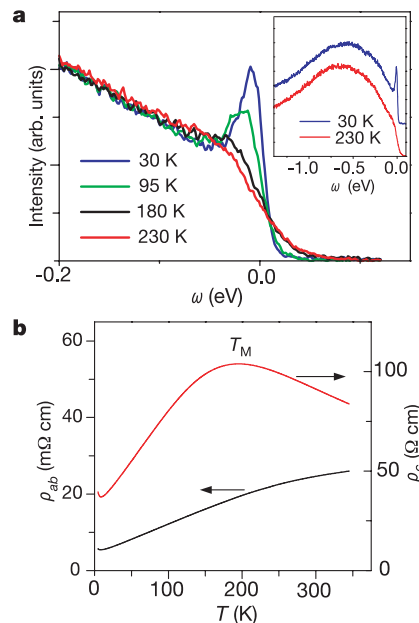


Figure 2 Correlation between the ARPES and transport in (Bi_{0.5}Pb_{0.5})₂Ba₃Co₂O₇. **a**, The changes in energy distribution curves (vertical cross-sections of the ARPES data shown in Fig. 1) (for $k = k_F$) with temperature. The inset shows the wide-range energy distribution curves. **b**, Transport data. The in-plane and the out-of-plane resistivities are measured on a sample from the same batch with a conventional four-probe technique.

$k_B T$ becomes comparable to that energy scale, the state loses coherence and the sharp peak diminishes. This is illustrated in panels Fig. 1b and c, where the spectral function is plotted on the same scale with the Fermi distribution for two given temperatures. With increasing temperature the spectrum loses its fine structure near the Fermi level. In Fig. 2 we show the temperature development of the spectral function in more detail, and correlate it with the crossover in transport. The energy distribution curves for $k \approx k_F$ are shown for several temperatures. The sharp peak both loses intensity and broadens with temperature, with the final result that between 180 K and 230 K the spectral function that was locally peaked at $\omega = 0$ at low temperature becomes a monotonically increasing function of energy at higher temperatures. The low-energy structure is lost, and only the broad hump characterizes the spectrum. These changes in the spectral function correlate with T_M (≈ 200 K) in c -axis resistivity.

The second system that we studied was NaCo_2O_4 . This material crystallizes in a layered hexagonal structure with a triangular Co lattice, octahedrally coordinated with O above and below the Co sheets. The Na ions are in the planes between the CoO_2 layers. The transport properties show many peculiarities⁴, including a c -axis crossover at $T_M \approx 180$ K (Fig. 3c). This system is a much better conductor than the previous one, with $\rho_{ab} \approx 300 \mu\Omega \text{ cm}$ and an anisotropy of ~ 40 at room temperature. The anisotropy increases at lower temperatures and saturates below ~ 120 K at a value of ~ 150 .

The spectral function of NaCo_2O_4 , shown in Fig. 3, shares common features with the previous system: a broad hump at high binding energy (~ 1.2 eV FWHM) and a sharp feature that crosses the Fermi level. At high temperatures, the sharp state again disappears and the low-energy part of the spectrum is a smooth, increasing function of energy. This transformation in the spectrum again correlates with the crossover in resistivity.

When the coherent quasiparticles form in the plane (in both materials), the c -axis transport becomes metallic and the system

behaves as an anisotropic 3D metal. It is not clear, however, whether the in-plane coherence is a consequence or the cause of the dimensional crossover. One possibility is that the coherence occurs and the third dimension develops when the temperature becomes smaller than the effective energy scale for c -axis hopping. With all three dimensions 'visible' ($k_B T \ll t_i$, $i = x, y, z$), the low-energy excitations are quasiparticles. This is in line with the reasoning characterizing quasi-1D metals, where the 1D to 3D crossover is dictated by the finite inter-chain ($t_\perp$) hopping integral¹⁰. In an alternative picture, sufficiently coherent ($\Gamma(T) \ll Z t_z$) quasiparticles must first be formed in the planes to induce or allow coherent c -axis transport between the planes.

The crossover from coherent to incoherent excitations and a maximum in resistivity at a finite temperature is reminiscent of Kondo behaviour in heavy-fermion systems¹¹. The latter crossover is related to the appearance of coherent, strongly renormalized (effective mass $m^* \approx 100\text{--}1,000 m_e$, where m_e is the bare mass) quasiparticle states below the temperature at which the f -moments order, as recently observed in a photoemission study¹². The systems studied here are, on the other hand, close to a Mott–Hubbard type of metal-to-insulator transition, and in one of them, $(\text{Bi}_{1-x}\text{Pb}_x)_2\text{Ba}_2\text{Co}_2\text{O}_y$, the transition is experimentally realized by changing the carrier concentration by lead doping⁵. Recent theories of the metal-to-insulator transition^{13–15} suggest that the same physics responsible for Kondo behaviour also shapes the physical properties near the metal-to-insulator transition. In the metallic regime, these theories predict a sharp peak in the density of states near the Fermi level, in addition to separated lower and upper Hubbard bands. The sharp peak corresponds to a strongly renormalized quasiparticle band whose effective mass diverges as the ratio of the onsite Coulomb energy to intersite hopping energy, U/t , increases to some critical value at the metal-to-insulator transition. This state is responsible for the Fermi-liquid behaviour at low temperatures. However, it disappears above some characteristic temperature, and the spectral function, controlled by the high-energy physics (U), becomes incoherent. The transport is no longer governed by coherent quasiparticles, but by collective excitations with resistivities going well beyond the Mott–Ioffe–Regel limit into a regime where the quasiparticle mean free path would be smaller than the interatomic distances¹⁶.

Depending on the ratio U/t , the incoherent response may even acquire a form characterizing the insulating side of the metal-to-insulator transition. It is important to note that although these models capture the incoherence–coherence transition, they are isotropic and therefore inappropriate for low-dimensional systems. In anisotropic materials, the different directions may be affected differently by electronic correlations. In the case of 1D Hubbard chains, for example, a small inter-chain hopping may induce deconfinement and Luttinger (1D, incoherent physics) to Fermi liquid (3D, coherent excitations) crossover¹⁷. For layered metals, equation (1) formally means that once the coherent part of the in-plane Green's function has disappeared ($Z \rightarrow 0$), $\rho_c(T)$ is no longer bound by the in-plane transport. In other words, if the response is governed by collective excitations instead of quasiparticles, it may take different forms in different directions.

In other systems, a similar dimensional crossover to a coherent, 3D low-temperature state may be induced by different mechanisms. For example, the copper oxide high-temperature superconductors near optimal doping have metallic planes with an insulating inter-layer resistivity, again suggesting that the normal-state transport is a collective phenomenon⁸. Indeed, the spectral function shows the absence of sharp quasiparticle peaks in the normal state. However, these systems become (anisotropic) 3D superconductors and the quasiparticles reappear below the superconducting transition temperature, T_C , in analogy with the dimensional crossovers discussed above. The transition goes directly from 2D-like normal state into the quasiparticle-like 3D superconducting state, and cannot be

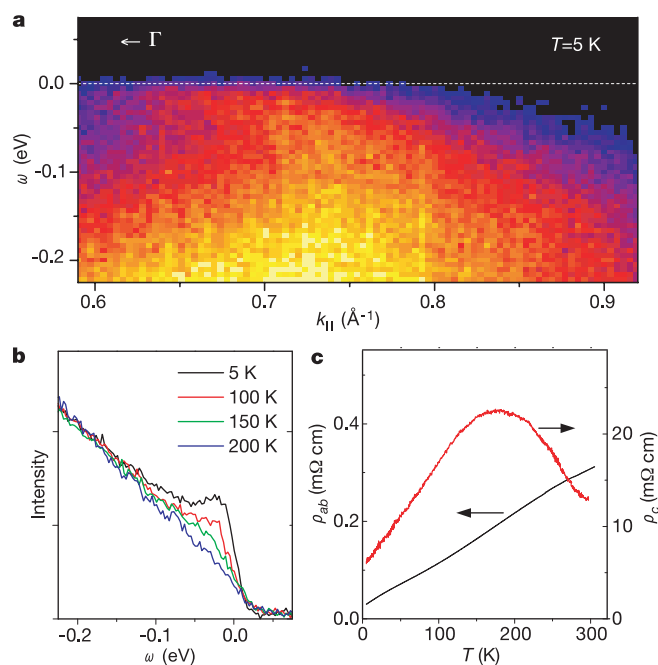


Figure 3 ARPES and d.c. transport in NaCo_2O_4 . **a**, Contour plot showing the ARPES intensity along the Γ – K line of the Brillouin zone at $T = 5$ K. The Fermi crossing is a point on a large hole-like Fermi surface ($k_F \approx 0.7 \text{ \AA}^{-1}$) centred at the Γ -point. The photon energy was 21.22 eV (He I radiation). The samples are flux grown. **b**, The temperature dependence of the low-energy region of energy distribution curves (for $k = k_F$). **c**, The in-plane and out-of-plane resistivities, measured on a sample from the same batch.

dictated by the single-particle tunnelling. In the stripe picture, for example, the normal state is considered in terms of the 1D physics within the charge stripes and the superconducting transition is viewed as a 1D to 3D transition, induced by the inter-stripe Josephson tunnelling of Cooper pairs that already exist within the stripes in the normal state¹⁸.

In an alternative model, a different behaviour for in-plane and out-of-plane transport could occur if both the interlayer hopping $t_{\perp}$, and the Fermi surface in the plane are anisotropic⁹. Indeed, it has been shown¹⁹ that in high-temperature superconductors, the scattering rate in the corners of the Brillouin zone that dominate c -axis hopping behaves differently from the scattering rate in the nodal regions²⁰ where $t_{\perp} = 0$. In addition, the corners of the Brillouin zone are affected by the normal-state pseudogap²¹ for $T_C < T < T^*$ (T^* is the pseudogap temperature) that blocks the normal-state c -axis transport. The inter-layer Josephson coupling then enables the transition into the 3D superconducting state without ever going through the 3D Fermi-liquid-like normal state. In the highly over-doped regime, the pseudogap disappears: the charges are then no longer confined to the planes²² ($\partial\rho_c/\partial T > 0$), even in the normal state. This transition into a 3D-like state, whether induced by single-particle tunnelling (normal state) or Josephson coupling (superconducting state), is accompanied by the appearance of a sharp peak in the spectral function²³, an indication that the 3D state approaches the Fermi liquid regime. The crossover may also proceed via the exchange coupling $J_{\perp}$, leading to an (anti)ferromagnetic 3D ground state. In copper oxides in the low-doping regime, there are indications²⁴ that different mechanisms may compete in bringing the system into the 3D ground state. $\square$

Received 19 November 2001; accepted 12 April 2002; doi:10.1038/nature00774.

- Mott, N. F. *Metal-insulator Transitions* (Taylor & Francis, London, 1990).
- J      , D. & Caron, L. G. *Low-dimensional Conductors and Superconductors* (Plenum, New York, 1987).
- Maeno, Y. *et al.* Superconductivity in a layered perovskite without copper. *Nature* **372**, 532–534 (1994).
- Terasaki, I., Sasago, Y. & Uchinokura, K. Large thermoelectric power in NaCoO₂ single crystals. *Phys. Rev. B* **56**, R12685–R12687 (1997).
- Loureiro, S. M. *et al.* Enhancement of metallic behaviour in bismuth cobaltates through lead doping. *Phys. Rev. B* **63**, 094109–1–094109–9 (2001).
- Tsukada, I. *et al.* Ferromagnetism and large negative magnetoresistance in Pb doped Bi-Sr-Co-O misfit-layer compound. Preprint cond-mat/0012395 at (<http://xxx.lanl.gov>) (2000).
- Mihaly, G., Kezsmarki, L., Zambrosky, F. & Forro, L. Hall effect and conduction anisotropy in the organic conductor (TMTSF)₂PF₆. *Phys. Rev. Lett.* **84**, 2670–2673 (2000).
- Anderson, P. W. *The Theory of Superconductivity in the High-T_c Cuprates* (Princeton Univ. Press, Princeton, 1997).
- Ioffe, L. B. & Millis, A. J. Zone-diagonal-dominated transport in high-T_c cuprates. *Phys. Rev. B* **58**, 11631–11637 (1998).
- Voit, J. One-dimensional Fermi liquids. *Rep. Prog. Phys.* **57**, 977–1116 (1994).
- Fisk, Z. *et al.* Heavy-electron metals: New highly correlated states of matter. *Science* **239**, 33–42 (1988).
- Reinert, F. *et al.* Temperature dependence of the Kondo resonance and its satellites in CeCu₂Si₂. *Phys. Rev. Lett.* **87**, 106401–1–106401–4 (2001).
- Pruschke, T., Cox, D. L. & Jarrell, H. Hubbard model at infinite dimensions: Thermodynamic and transport properties. *Phys. Rev. B* **47**, 3553–3565 (1993).
- Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68**, 13–125 (1996).
- Merino, J. & McKenzie, R. H. Transport properties of strongly correlated metals: A dynamical mean-field approach. *Phys. Rev. B* **61**, 7996–8008 (2000).
- Ioffe, A. F. & Regel, A. R. Non-crystalline, amorphous and liquid electronic semiconductors. *Prog. Semicond.* **4**, 237–291 (1960).
- Biermann, S., Georges, A., Lichtenstein, A. & Gianmarchi, T. Deconfinement transition and Luttinger to Fermi liquid crossover in quasi one-dimensional systems. Preprint cond-mat/0107633 at (<http://xxx.lanl.gov>) (2001).
- Carlson, E. W., Orgad, D., Kivelson, S. A. & Emery, V. J. Dimensional crossover in quasi-one-dimensional and high-T_c superconductors. *Phys. Rev. B* **62**, 3422–3437 (2000).
- Valla, T. *et al.* Temperature dependent scattering rates at the Fermi surface of optimally doped Bi₂Sr₂CaCu₂O_{8+ }. *Phys. Rev. Lett.* **85**, 828–831 (2000).
- Valla, T. *et al.* Evidence for quantum critical behaviour in the optimally doped cuprate Bi₂Sr₂CaCu₂O_{8+ }. *Science* **285**, 2110–2113 (1999).
- Ding, H. *et al.* Spectroscopic evidence for a pseudogap in the normal state of underdoped high-T_c superconductors. *Nature* **382**, 51–54 (1996).
- Watanabe, T., Fujii, T. & Matsuda, A. Pseudogap in Bi₂Sr₂CaCu₂O_{8+ } studied by measuring anisotropic susceptibilities and out-of-plane transport. *Phys. Rev. Lett.* **84**, 5849–5852 (2000).
- Yusuf, Z. *et al.* Quasiparticle liquid in the highly overdoped Bi₂Sr₂CaCu₂O_{8+ }. Preprint cond-mat/0104367 at (<http://xxx.lanl.gov>) (2001).
- Lavrov, A. N., Kameeva, M. Y. & Kozeeva, L. P. Normal-state resistivity anisotropy in underdoped RBa₂Cu₃O_{6+x} crystals. *Phys. Rev. Lett.* **81**, 5636–5639 (1998).

Acknowledgements

We thank A. Tselik, R. Werner, S. A. Kivelson and A. V. Fedorov for discussions. The work at BNL and the National Synchrotron Light Source where the experiments were carried out was supported by the US Department of Energy. B.W. thanks the A.P. Sloan foundation for a Research Fellowship.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.V. (e-mail: valla@bnl.gov).

Measurement of a confinement induced neutron phase

H. Rauch*, H. Lemmel*, M. Baron*† & R. Loidl*†

*Atominstytut der   sterreichischen Universit  ten, A-1020 Wien, Austria

†Institut Laue-Langevin, BP 156, F-38042 Grenoble, France

Particle physicists see neutrons as tiny massive particles with a confinement radius of about 0.7 fm and a distinct internal quark–gluon structure. In quantum mechanics, neutrons are described by wave packets whose spatial extent may become ten orders of magnitude larger than the confinement radius, and can even reach macroscopic dimensions, depending on the degree of monochromaticity. For neutrons passing through narrow slits, it has been predicted^{1,2} that quantization of the transverse momentum component changes the longitudinal momentum component, resulting in a phase shift that should be measurable using interferometric methods³. Here we use neutron interferometry to measure the phase shift arising from lateral confinement of a neutron beam passing through a narrow slit system. The phase shift arises mainly from neutrons whose classical trajectories do not touch the walls of the slits. In this respect, the non-locality of quantum physics is apparent.

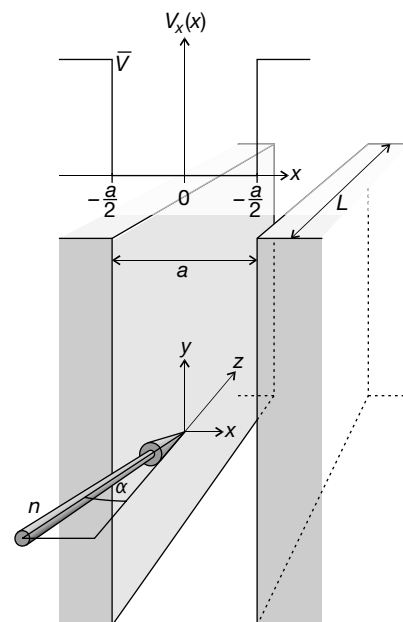


Figure 1 Sketch of the slit structure and of the related neutron–wall interaction potential. Arrow denotes neutron beam n ; $\bar{V}$, neutron optical potential; a , width of potential; L , length of the slit.

A spatial confinement can change the physical properties of a trapped system in a similar sense to that in which frequent measurements of the occupied state can do so⁴⁻⁷. This can be understood from the limitation in phase space caused by the spatial or temporal confinement. It has also been observed that the atomic dimensions and the van der Waal-like interaction of atoms and molecules cause, in the case of matter-wave diffraction, deviations from classical optical phenomena^{8,9}. Similar effects exist for electrons during ballistic motion through specific semiconducting heterostructures¹⁰ and for light passing through standing laser fields^{11,12}.

The special feature of the neutron lies in its extremely small dimension and small interaction range ($<10^{-15}$ m). Measurements for the lateral confinement effect with laser beams¹³ can be considered as test measurements for the experiments described here. The recently observed quantization of ultra-cold neutrons in the Earth's gravitational field shows a confinement effect above a flat surface¹⁴.

For a proper description of the behaviour of neutrons inside narrow slits the wave packet formalism for a neutron beam has to be used. Quantum mechanics permits the calculation of the wavefunction for the neutron quantum states inside such structures. The solution of the stationary Schrödinger equation for a rectangular, infinitely high potential shows quantized states with transverse momentum components $k_{n\perp} = n\pi/a$, where a denotes the width of the potential and $n = 1, 2, 3, \dots$. In this experiment, the potential is finite and the height is given by the neutron optical potential, $\bar{V} = 2\pi\hbar^2 b_c N/m$, which is determined by the coherent scattering length b_c , the particle density N of the wall material and the mass of the neutron, m . For silicon as wall material and a slit width of $a = 22.1 \mu\text{m}$, as used in the experiment, this leads to energy levels of the order of peV. In this case, 360 levels lie within the potential (bound states) and a series of free states follow above $\bar{V}$.

Figure 1 shows the slit system together with the related rectangular potential and indicates how the beam divergence has to be taken into account. Owing to intensity requirements, the experiments have to be performed with a multi-slit system and a more detailed theoretical analysis is required. The degree of excitation of the levels depend on the slit width and on the angle of incidence α and is denoted by weight factors $p(\alpha)$. For example, a completely parallel beam (with $\alpha = 0^\circ$) excites the symmetric levels $E_0 = 0.4172$ peV with $p_0(0) = 0.406$, $E_2 = 1.669$ peV with $p_2(0) = 0.045$ and $E_4 = 6.676$ peV with $p_4(0) = 0.016$. In contrast, a beam component which deviates by an angle $\alpha = 0.005^\circ$ excites distinct energy levels around $E_{19} = 166.89$ peV and so on. Behind the slits each incident beam component excites two beams, one in the same direction (forward) and another one which is totally reflected and directed in

a slightly different direction. In an interference experiment with a symmetric reference beam only the beam component in the forward direction contributes to the interference pattern. The resulting phase factor connecting the incident and outgoing neutron wavefunction ψ_{in} and ψ_{out} for each component is therefore given as:

$$S(\alpha) = \sum_{n=0}^N p_n(\alpha) e^{i\Delta\phi_n} \quad (1)$$

for $\psi_{\text{out}} = S(\alpha)\psi_{\text{in}}$, where $\Delta\phi_n = L(k_{\parallel,n} - k_{\parallel,\text{in}})$ denotes the phase shift due to level n . $k_{\parallel,n}$ and $k_{\parallel,\text{in}}$ indicate the parallel momentum inside and in front of the slit system, respectively:

$$k_{\parallel,n} = \sqrt{2m(E_{\text{in}} - E_n)/\hbar^2} \quad (2)$$

where E_{in} denotes the energy of the incident beam ($E_{\text{in}} \gg E_n$).

The variation of the slit width $\Delta a/a$ smears out the contribution to the neutron phase shift from the higher levels and the contributions from the free levels are smeared out owing to both the variation of the slit length and the wavelength distribution. The remaining predicted phase shift for $\Delta a/a = 0.14$, which is the value measured for the slit system used, is shown in Fig. 2. It is mainly the nearly parallel beam components (low-lying levels) that contribute to a measurable phase shift and a measurable contrast. The calculated phase shift ($\sim 2.5^\circ$) for the parallel beam components is rather insensitive to the variation of the slit width $\Delta a/a$, indicating that the contributing beam components classically do not touch the slit walls at all. The expected phase shift varies approximately with $1/a^2$ as a function of the slit width. For values $\alpha > 0.063^\circ$ each classical neutron trajectory touches, at least once, a slit wall and the resulting phase shift oscillates strongly, indicating a transition to a guided wave behaviour. The beam divergence in the horizontal direction was 0.35° which is about four times the critical angle for total reflection. Thus we find beam components that do not touch

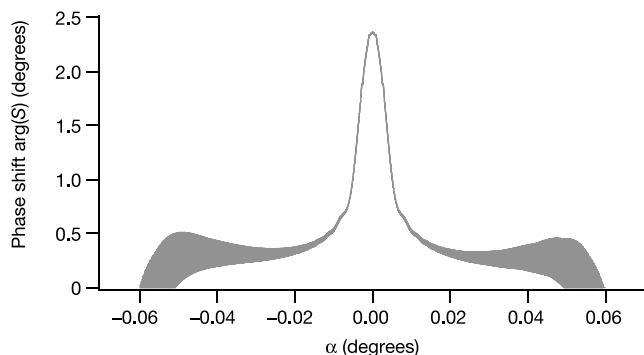


Figure 2 Expected phase shift as a function of the angle α of the incident beam component. In this calculation, the slit structure, the divergency of the beam and the variation of the slit width are taken into account. $\text{Arg}(S)$, the argument of the phase factor $S(\alpha)$.

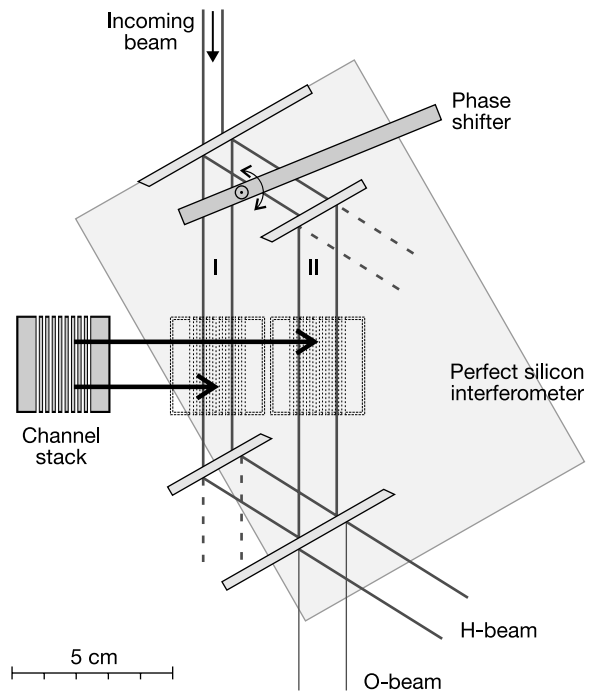


Figure 3 Sketch of the experimental set-up. The neutron beam emerges from a thermal moderator of the high flux reactor, is guided through a 25-m-long super mirror neutron guide, and is monochromatized by Bragg diffraction from a perfect silicon crystal. The skew symmetric perfect crystal neutron interferometer provides two widely separated coherent beams (I and II) which are used for phase shift measurements. An auxiliary aluminium phase shifter is rotated within the beams to produce the related interference pattern. (O-, H-beam denotes forward and deviated beam, respectively.)

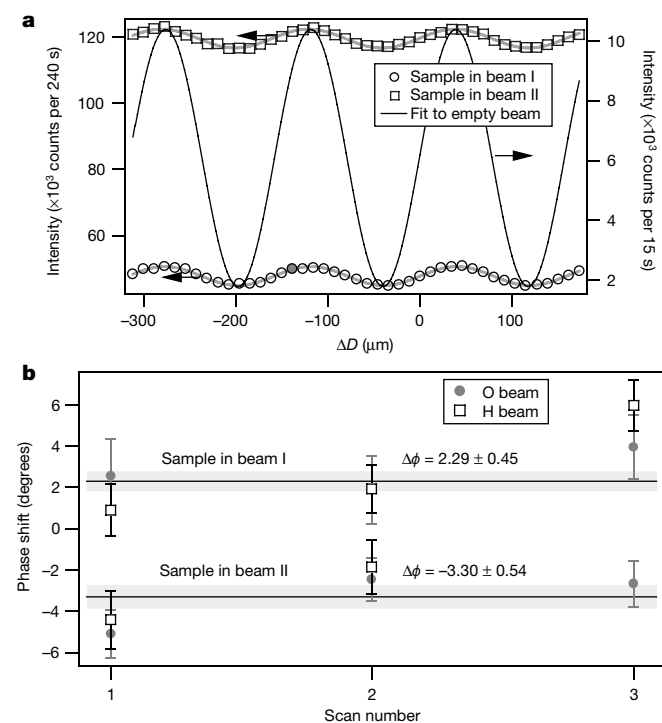


Figure 4 Typical interference pattern with and without the phase-shifting slit system (**a**) and the collection of the results of various scans (**b**). ΔD denotes the path length difference of beam I and II within the phase shifter.

the walls but which do contribute to the expected phase shift; we also find beam components that are totally reflected from the walls or that pass through the walls, neither of which contribute to a measurable contrast owing to smearing effects.

The measurements were performed at the neutron interferometer set-up S18 at the Institut Laue-Langevin in Grenoble. A skew symmetric interferometer has been used and adjusted to a wavelength of $\lambda = 1.89(3)$ Å, which corresponds to an incident energy of $E_{\text{in}} = 0.0278(8)$ eV. The slit system has been put into beams I and II, and outside both beams (Fig. 3). This sweep procedure permits the measurement of small phase shifts at low contrast. The slit system consists of 186 slits with a width of 22.1 μm and the same wall thickness and a length of 20 mm. Silicon wafers were used as wall material and spacers. After careful adjustment to the parallel position ($\alpha \cong 0^\circ$) various scans have been made where Fig. 4 shows a typical result and the summary result of all scans. As expected, the opposite phase shift becomes visible for both beams and this is a good indication that spurious phases have been excluded. The measured value of the confinement-induced phase shift is $\Delta\phi = 2.8(4)^\circ$, which is in rather good agreement with the expected value ($\Delta\phi \cong 2.5^\circ$).

The wavefunctions within the slits show an interesting pattern (wavefunction carpets) which could be exploited in more detail in future. When the slits are oriented horizontally, a triangular-shaped potential due to the gravity effect occurs that changes mainly the low-lying levels from $E_0 = 0.417$ peV to $E_{0'} = 0.2904$ peV or from $E_1 = 1.669$ peV to $E_{1'} = 1.702$ peV (refs 15, 16). For such experiments the interferometer would have to be rotated to the horizontal direction as well because the large wave packet extension needed for these experiments exists only in the direction normal to the reflecting crystal planes of the interferometer^{17,18}. A different situation occurs when absorbing wall materials are used. Other effects may occur when a time-dependent interaction (for example, vibrating walls) can cause transitions between different energy levels. □

Received 22 January; accepted 17 April 2002; doi:10.1038/nature00773.

1. Lévy-Leblond, J. M. A geometrical quantum phase effect. *Phys. Lett. A* **125**, 441–442 (1987).
2. Greenberger, D. M. A new non-local effect in quantum mechanics. *Physica B* **151**, 374–377 (1988).
3. Rauch, H. & Werner, S. A. *Neutron Interferometry* (Clarendon, Oxford, 2000).
4. Casimir, H. B. G. & Polder, D. Influence of retardation on the London-van der Waals forces. *Phys. Rev.* **73**, 360–372 (1948).
5. Haroche, S. & Raimond, J. M. Cavity quantum electrodynamics. *Sci. Am.* **268**, 26–33 (1993).
6. Fonda, L., Ghirardi, G. C., Rimini, A. & Weber, T. On the quantum foundations of the experimental decay law. *Nuovo Cimento A* **15**, 689–704 (1973).
7. Misra, B. & Sudarshan, E. C. G. The Zeno's paradox in quantum theory. *J. Math. Phys.* **18**, 756–763 (1977).
8. Grisenti, R. E. *et al.* Determination of atom-surface van der Waals potentials from transmission diffraction intensities. *Phys. Rev. Lett.* **83**, 1755–1758 (1999).
9. Hegerfeldt, G. C. & Koehler, T. Deviations from classical optics in matter diffraction and determination of the size of weakly bound molecules. *Phys. Rev. A* **61**, 023606–1–023606–10 (2000).
10. Ferry, D. K., Grubini, H., Jacobini, C. & Jauho, A. J. (eds) *Quantum Transport in Ultrasmall Devices* (Plenum, New York, 1995).
11. Salomon, C., Dalibard, J., Aspect, A., Metcalf, H. & Cohen-Tannoudji, C. Channeling atoms in a laser standing wave. *Phys. Rev. Lett.* **59**, 1659–1662 (1987).
12. Keller, C. *et al.* Adiabatic following in standing-wave diffraction of atoms. *Appl. Phys. B* **69**, 303–309 (1999).
13. Allman, B. E., Cimmino, A., Griffin, S. L. & Klein, A. G. Quantum phase shift caused by spatial confinement. *Found. Phys.* **29**, 325–332 (1999).
14. Nesvizhevsky, V. V. *et al.* Quantum states of neutrons in the Earth's gravitational field. *Nature* **415**, 297–299 (2002).
15. Flügge, S. *Practical Quantum Mechanics* (Springer, Berlin, 1971).
16. Pokotilovski & Yu, N. Quantum phase shift of spatially confined de Broglie waves in a gravitational field. *Phys. Lett. A* **248**, 114–116 (1998).
17. Rauch, H., Wölwitsch, H., Kaiser, H., Clothier, R. & Werner, S. A. Measurement and characterization of the three-dimensional coherence function in neutron interferometry. *Phys. Rev. A* **53**, 902–908 (1996).
18. Rauch, H. & Summhammer, J. Neutron interferometer absorption experiments in the quantum limit. *Phys. Rev. A* **46**, 7284–7287 (1992).

Acknowledgements

This work was supported by the Austrian Science Foundation and a TMR-Network of the European Union EU. Useful discussions with J. Summhammer and D. Petrascheck are gratefully acknowledged.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.R. (e-mail: rauch@ati.ac.at).

Marine aerosol formation from biogenic iodine emissions

Colin D. O'Dowd^{*,†}, Jose L. Jimenez[‡], Roya Bahreini[‡], Richard C. Flagan[‡], John H. Seinfeld[‡], Kaarle Hämeri[§], Liisa Pirjola^{||}, Markku Kulmala[†], S. Gerard Jennings^{*} & Thorsten Hoffmann[¶]

^{*} Department of Physics, National University of Ireland, Galway, Ireland

[†] Division of Atmospheric Sciences, Department of Physical Sciences, P.O. Box 64, FIN-00014, University of Helsinki, Helsinki, Finland

[‡] Departments of Environmental Science and Engineering and Chemical Engineering, California Institute of Technology, 210-41, 1200 E. California Boulevard, Pasadena, USA

[§] Finnish Institute for Occupational Health, Topeliuksenkatu 41a, FIN-00250, Helsinki, Finland

^{||} Helsinki Polytechnic Technology, PL 4020, FIN-00099, Helsinki, Finland

[¶] Institute of Spectrochemistry and Applied Spectroscopy, 44139 Dortmund, Germany

The formation of marine aerosols and cloud condensation nuclei—from which marine clouds originate—depends ultimately on the availability of new, nanometre-scale particles in the marine boundary layer. Because marine aerosols and clouds scatter incoming radiation and contribute a cooling effect to the Earth's radiation budget¹, new particle production is important

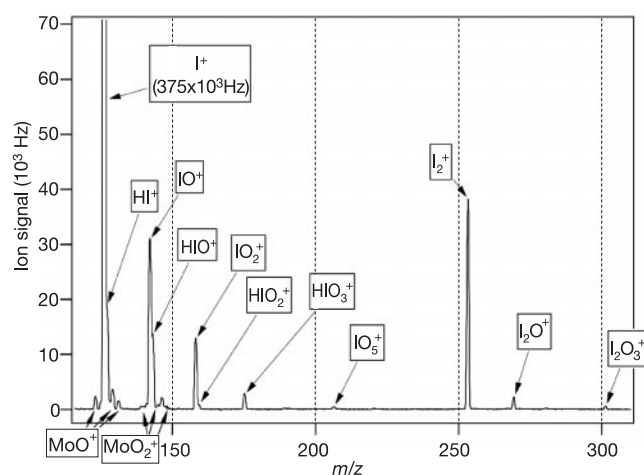


Figure 1 Observed aerosol mass spectrum from chamber experiments as measured by the aerosol mass spectrometer. The aerosol resulted from the photo-oxidation of CH_2I_2 in the presence of O_3 in pure laboratory air. The observed peaks are consistent with particles composed of iodine oxides and/or iodine oxyacids, and possibly also water. The molybdenum oxide peaks observed in the spectrum are probably due to the oxidation of the molybdenum particle vaporizer of the spectrometer by iodine oxyacids contained in the sampled particles.

in climate regulation. It has been suggested that sulphuric acid—derived from the oxidation of dimethyl sulphide—is responsible for the production of marine aerosols and cloud condensation nuclei. It was accordingly proposed that algae producing dimethyl sulphide play a role in climate regulation², but this has been difficult to prove and, consequently, the processes controlling marine particle formation remains largely undetermined^{3,4}. Here, using smog chamber experiments under coastal atmospheric conditions, we demonstrate that new particles can form from condensable iodine-containing vapours, which are the photolysis products of biogenic iodocarbons emitted from marine algae. Moreover, we illustrate, using aerosol formation models, that concentrations of condensable iodine-containing vapours over the open ocean are sufficient to influence marine particle formation. We suggest therefore that marine iodocarbon emissions have a potentially significant effect on global radiative forcing.

There are two steps required for the production of new particles by gas-to-particle conversion processes (secondary particle formation). First, a critical embryo, or thermodynamically stable cluster of the order of 1 nm in size, must be formed from the coalescence of precursor vapours (homogeneous nucleation)⁵. Second, these stable clusters must grow rapidly enough, by coagulation and/or condensation, to quasi-stable particle sizes of 3–4 nm (particle formation) before they collide by diffusion with larger, pre-existing, particles⁶ and are subsequently captured (diffusion decreases as $1/D^2$, where D is the particle diameter). Once new particles reach sizes of a few nanometres, there is a net increase in the particle number concentration.

Recent simulations of ternary H_2SO_4 – H_2O – NH_3 nucleation⁷ have shown that under typical atmospheric concentrations of H_2SO_4 , nucleation of thermodynamically stable clusters can readily occur; however, production of new particles via growth by vapour condensation is unlikely, owing to insufficient supply of H_2SO_4 . Although H_2SO_4 could be responsible for the production of thermodynamically stable clusters, other vapours are required for particle production. The most comprehensive studies⁸ into marine aerosol nucleation have focused on coastal environments, where regular and intense particle production events are encountered owing to the higher intensity of biological processes in coastal areas. Such particle production is independent of peak H_2SO_4 concentrations⁹, suggesting a hitherto unknown marine particle production mechanism. Experimental measurements on new particle composition have ruled out a chemical composition dominated by H_2SO_4 (ref. 10), and identified iodine in these particles¹¹. Consequently, the production of particles in the coastal zone by condensable iodine vapours (CIVs) resulting from the photolysis of biogenic CH_2I_2 has been proposed⁸.

To test this hypothesis, we conducted experiments (see Methods) in the Caltech smog chamber¹² into the particle production ability of CH_2I_2 undergoing photolysis in the presence of ozone, and we present results confirming that this is a viable mechanism for particle production down to atmospheric concentrations¹³. No particles were formed unless CH_2I_2 , O_3 and ultraviolet radiation were simultaneously present in the chamber. When ultraviolet radiation was introduced after CH_2I_2 and O_3 had been mixed throughout the chamber, rapid homogeneous nucleation occurred, followed by condensation and coagulation growth. This behaviour was observed for all experiments for CH_2I_2 levels from 50 parts per billion (50 p.p.b.) down to the lowest CH_2I_2 concentration achievable in this chamber of 0.015 p.p.b. ($\sim 4 \times 10^8$ molecules cm^{-3}).

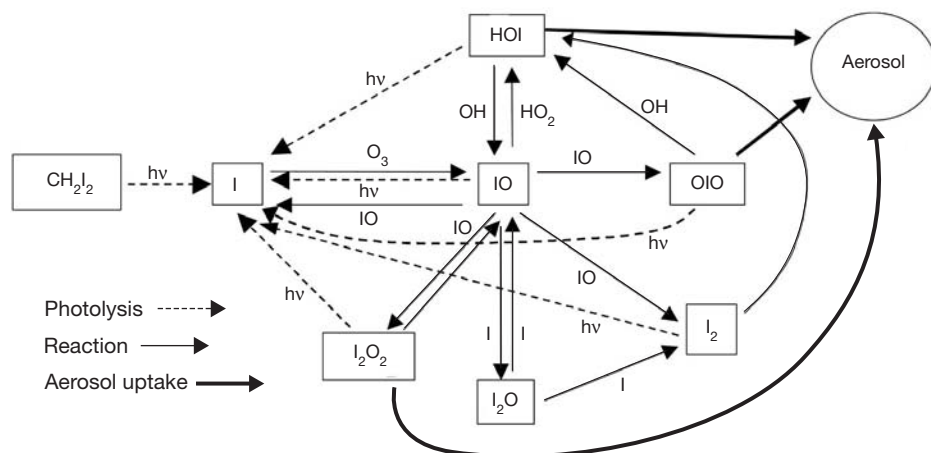


Figure 2 Chemical pathway from CH_2I_2 to aerosol production, on the basis of the current state of knowledge of the gas-phase chemistry. Quantitative rates are given in refs 18 and 30.

Decreasing the CH_2I_2 concentration did not change the nucleation or growth phenomenology, but resulted in longer timescales for both processes.

Particles formed under dry conditions possessed a fractal morphology, whereas particles formed under humid conditions (relative humidity $\sim 65\%$) were more compact and dense. No other significant differences in the nucleation or growth behaviour as a function of humidity was observed. Reducing the intensity of the ultraviolet radiation slowed the particle formation processes without changing their qualitative features. Analysis of the resultant aerosol chemical composition, using an aerosol mass spectrometer, revealed the following mass/charge (m/z) peaks: OH^+ , H_2O^+ , I^+ , HI^+ , IO^+ , HIO^+ , IO_2^+ , HIO_2^+ , HIO_3^+ , IO_3^+ , I_2^+ , I_2O^+ and I_2O_3^+ (Fig. 1). These peaks are consistent with particles composed of iodine oxides and/or iodine oxyacids, and possibly also water. Particle composition did not change with particle size or with time in the experiment.

The chemical mechanisms associated with the particle production is presented in Fig. 2. There are three possible pathways to form CIVs: (1) CH_2I_2 is rapidly photolysed, releasing an I atom which reacts with O_3 to produce IO . IO in turn reacts with itself to produce OIO , which can nucleate to produce I_2O_4 and higher-order iodine oxide polymers; (2) IO can also self-react to produce I_2O_2 , which can participate in aerosol formation; (3) IO can react with HO_2 to produce condensable HOI . At the lowest concentrations of less than 20–200 parts per trillion (p.p.t.) CH_2I_2 , HOI is the dominant CIV, whereas iodine oxides dominate at higher concen-

trations. These results show that nucleation and condensation of CIVs occurs under atmospheric coastal conditions, and that this mechanism must be considered viable in terms of contributing to the marine aerosol population.

Estimates^{14–17} of the global flux of iodine from macroalgae are of the order of 10^7 – 10^8 g yr^{-1} , but this source is limited to coastal zones. Microalgae (phytoplankton), on the other hand, are found all over the ocean, and are estimated to contribute 10^9 – 10^{10} g I yr^{-1} . The global ocean source, however, is estimated^{14–17} to be 10^{11} – 10^{12} g I yr^{-1} , but the main contributor to this source of iodine, in addition to micro and macro algae, is currently unidentified, although there is strong evidence that bacteria can synthesize CH_3I , which could account for the missing source¹⁷. Nevertheless, taking a global marine source of 10^{12} g I yr^{-1} and a globally averaged marine surface-mixed-layer height of 300 m, the source rate of iodine atoms is estimated at 1.4×10^3 $\text{atoms cm}^{-3} \text{ s}^{-1}$. This open-ocean source of iodine atoms compares favourably to the recent estimate of 1×10^4 atoms cm^{-3} required to explain the observed concentration of IO and OIO (up to 1.6×10^8 cm^{-3}) over the open ocean^{18,19}.

We simulate, using a marine aerosol model (see Methods), the production of new particles resulting from condensation of CIVs onto thermodynamically stable cluster embryos. The simulation (Fig. 3) is started just before sunrise, and as dimethylsulphide is photochemically destroyed, H_2SO_4 production also increases, leading to a peak concentration of 2.5×10^6 molecules cm^{-3} . At these typical H_2SO_4 concentrations, the concentration of thermodynamically stable clusters reaches a maximum of $\sim 5,000$ cm^{-3} , resulting from a peak nucleation rate of 4 $\text{cm}^{-3} \text{ s}^{-1}$, but barely-detectable new particle formation results in sizes larger than 3 nm. When an additional CIV source (Q) of 10^3 $\text{cm}^{-3} \text{ s}^{-1}$ is included, slight particle production results after a few hours, resulting from an enhancement in condensable vapours of 0.5×10^6 cm^{-3} . Increasing Q to 5×10^3 , 10×10^3 and 25×10^3 $\text{molecules cm}^{-3} \text{ s}^{-1}$ leads to a CIV concentration of respectively 2.5×10^6 , 5×10^6 and 12×10^6 molecules cm^{-3} , and increases the particle concentration from 518 cm^{-3} to 555 cm^{-3} , 614 cm^{-3} and 804 cm^{-3} , respectively. Clearly, increasing the source rate of CIVs significantly increases the production rate of new particles.

The initial and final aerosol size distributions are given in Fig. 4 for all simulations. As the CIV vapour concentration increases, the final concentration of the nucleation mode increases, as does the

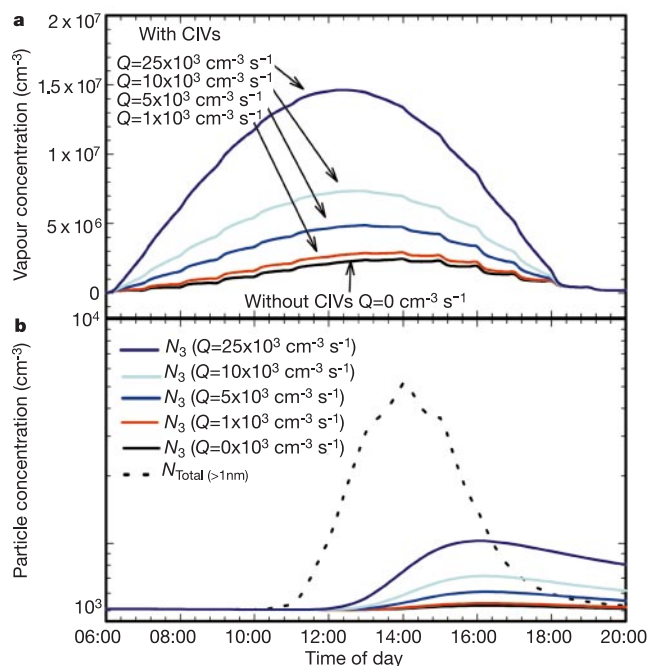


Figure 3 Model simulations of condensable vapour and aerosol concentrations in the marine boundary layer. **a**, Combined concentration of H_2SO_4 and condensable iodine vapours (CIVs) for CIV source rates (Q) of 0, 10^3 , 5×10^3 , 10×10^3 and 25×10^3 $\text{molecules cm}^{-3} \text{ s}^{-1}$ and a dimethylsulphide concentration of 20 p.p.t. (5.6×10^8 molecules cm^{-3}). Oscillations in the vapour concentration occur owing to the variation in the pre-existing aerosol condensation sink as the aerosol surface area responds to changes in relative humidity during air parcel cycling through the boundary layer. **b**, Predicted particle concentration (cm^{-3}) at sizes larger than 1 nm (N_{total}), including thermodynamically stable clusters; and detectable particle sizes of 3 nm and greater (N_3) for CIV source rates of 0, 10^3 , 5×10^3 , 10×10^3 and 25×10^3 $\text{molecules cm}^{-3} \text{ s}^{-1}$.

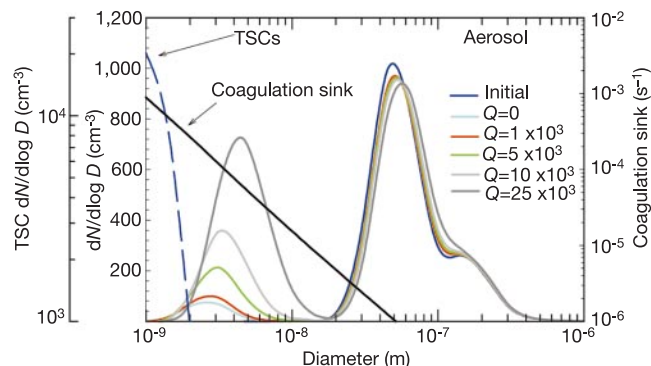


Figure 4 Initial (local time 06:00) and predicted final (local time 20:00) aerosol size distributions for CIV source rates (Q) of 0, 10^3 , 5×10^3 , 10×10^3 and 25×10^3 $\text{molecules cm}^{-3} \text{ s}^{-1}$. Also shown (note different scale) is the maximum concentration of thermodynamically stable clusters (TSCs) during the period of peak particle production (local time 14:00). The coagulation sink (the rate of loss of particles per second owing to coagulation with the pre-existing aerosol) as a function of particle size is superimposed on the figure. N , number of particles.

mean modal size (up to 6 nm). At a condensable vapour concentration of $2.4 \times 10^6 \text{ cm}^{-3}$ ($Q = 0$), the nucleation mode comprising thermodynamically stable clusters, initially 1 nm in size, cannot grow into sizes larger than 3 nm (more diameter) because the growth rate is slow (0.2 nm h^{-1}) relative to the coagulation sink²⁰ ($2 \times 10^{-3} \text{ s}^{-1}$ for a 1-nm particle, decreasing to $2 \times 10^{-4} \text{ s}^{-1}$ for a 3-nm particle). At a vapour concentration of $14 \times 10^6 \text{ cm}^{-3}$ ($Q = 25 \times 10^3 \text{ molecules cm}^{-3} \text{ s}^{-1}$), the growth rate is significantly faster (1.1 nm h^{-1}). Therefore, it takes one-fifth of the time to reach sizes of 3 nm where particles can survive 10 times longer. Similarly, as the particle grows to 6 nm, its loss rate is reduced even further to $4 \times 10^{-5} \text{ s}^{-1}$. In essence, this additional vapour source carries the nuclei over a coagulation loss barrier, resulting in a higher probability of the new particles surviving. If dimer nucleation of CIVs is considered²¹, nucleation rates are predicted to increase to $10^2 \text{ cm}^{-3} \text{ s}^{-1}$ for a CIV concentration of 10^6 cm^{-3} , increasing further to $10^4 \text{ cm}^{-3} \text{ s}^{-1}$ for a CIV concentration of 10^7 cm^{-3} , so CIVs are likely to also contribute to the nucleation process.

Overcoming the coagulation loss barrier essentially increases the number and lifetime of condensation sites for production of cloud condensation nuclei (CCN) during the next days' cycle. Typically, the CCN concentration for clean marine air is $50\text{--}100 \text{ cm}^{-3}$, accounting for 10–20% of the aerosol population. The increase in the net input of particles into the aerosol population is likely to result in a similar percentage increase in available CCN (corresponding to ~10% of the new particles surviving to CCN sizes). Therefore, changes in the source rate of biogenic CIVs associated with global climate change are likely to significantly influence the CCN population. In fact, studies into the emissions of iodocarbons from marine biota have shown that emissions can increase by up to 5 times as a result of changes in environmental conditions associated with global change¹⁴. Therefore, increasing the source rate of CIVs fivefold will result in an increase in marine aerosol and CCN concentrations of the order of 20–60%. Changes in cloud albedo resulting from changes in CCN concentrations of this magnitude can lead to an increase in global radiative forcing similar in magnitude, but opposite in sign, to the forcing induced by greenhouse gases²².

Although in some regions there is evidence for marine particle production from organic species²³, the proposed contribution of CIVs to the marine aerosol is most strongly supported by field measurements of marine aerosol composition^{24–28}: iodine enrichment factors as high as 1,000 relative to bulk sea water have been found in marine aerosols globally, reaching factors as high as 33,000 in coastal zones. The present work confirms an aerosol nucleation mechanism involving CIVs in coastal regions, and suggests that CIVs enhance the production of particles in conjunction with dimethylsulphide-derived H_2SO_2 nucleation over remote oceans. CIVs could also contribute to the remote-ocean nucleation—but to confirm this, more theoretical and experimental advances into understanding the atmospheric iodine cycle need to be achieved. □

Methods

Laboratory experiments

Aerosol formation from the photolysis of CH_2I_2 in the presence of O_3 was investigated in the Caltech indoor chamber. This facility has been described previously¹², and consists of dual, 28-m^3 chambers illuminated by 300 1.22-m fluorescent lights which are used to simulate the ultraviolet and near-ultraviolet regions of ambient sunlight. The ultraviolet light intensity in the chamber was measured with a portable spectroradiometer (LI-1800, LI-COR, Lincoln), and was about 1.2 times the maximum intensity reaching the Earth's surface with a zenith angle of 0° , except for one low radiation experiment which was performed with one-quarter of the full intensity. The reaction chamber was filled with particle-free dry air at 20°C that had been scrubbed of both volatile organic compounds and NO_x . CH_2I_2 (99% purity, Sigma-Aldrich) was dissolved in hexafluorobenzene (99% purity, Sigma-Aldrich), and the solution was evaporated into clean dry carrier air and injected into the chamber. After O_3 injection and allowing sufficient time for reactant mixing, the experiment was initiated by switching on the ultraviolet lights.

The evolution of the particle number concentration, number distribution, and hygroscopicity was monitored by condensation particle counters, differential mobility

analysers, and a tandem differential mobility analyser, respectively, as described¹². CH_2I_2 concentrations of 0.015 p.p.b. resulted in particle growth up to 30 nm (mode diameter), whereas for concentrations of 50 p.p.b., the maximum mode size reached was 200 nm. Particle composition was monitored using an Aerodyne aerosol mass spectrometer²⁹. This spectrometer uses an aerodynamic lens inlet to focus the sampled particles into a narrow beam, which is then introduced into a high-vacuum chamber while the air is differentially pumped. The particles are vaporized on striking a heated, roughened molybdenum cartridge heater under high vacuum (10^{-7} torr) at about 600°C . The vaporized species are then ionized by electron impact (70-eV electrons), and the positive ions produced are mass analysed with a quadrupole mass spectrometer. Particle aerodynamic diameter is determined by particle time-of-flight. The spectrometer can produce two types of data: mass spectra of the species present in/on the aerosol, and mass-weighted size distributions versus aerodynamic diameter at a series of m/z peaks of the mass spectrometer. The spectrometer used in these experiments was able to detect particles between about 40 nm and $1 \mu\text{m}$, with a minimum detectable concentration of about $0.5 \mu\text{g m}^{-3}$; it was calibrated for particle size with polystyrene latex spheres (Duke Scientific, Palo Alto).

Aerosol modelling

We used AEROFOR⁷, a lagrangian-based 27-bin sectional box model which includes gas-phase chemistry and aerosol dynamics. The following processes are considered: (1) dimethylsulphide surface flux; (2) gas-phase oxidation of dimethylsulphide to SO_2/SO_3 ; (3) gas-phase oxidation of SO_2/SO_3 to H_2SO_4 ; (4) homogeneous nucleation of H_2SO_4 , H_2O and NH_3 ; (5) condensation of CIVs, methane sulphonic acid, H_2SO_4 and H_2O onto pre-existing particles; (6) direct reaction of SO_2 with particles; (7) brownian coagulation of particles; (8) dry deposition to the sea surface. Full details can be found in ref. 7 and references therein. Cloud or precipitation processes are not included, and neither is heterogeneous removal of SO_2 . The simulations are initialized with a typical marine aerosol concentration of 510 cm^{-3} , comprising an Aitken mode, accumulation mode, sea-salt jet drop mode and sea-salt film-drop mode, dimethylsulphide concentrations of 20 p.p.t., and NH_3 and SO_2 concentrations of 16 and 5 p.p.t., respectively. The air parcel undergoes adiabatic cycling in the boundary layer with relative humidity ranging from 70% at the surface to 99% at the top of the boundary layer.

Received 21 December 2001; accepted 16 April 2002; doi:10.1038/nature00775.

- Slingo, A. Sensitivity of the Earth's radiation budget to changes in low clouds. *Nature* **343**, 49–51 (1990).
- Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326**, 655–661 (1987).
- Katoshevski, D., Nenes, A. & Seinfeld, J. H. A study of processes that govern the maintenance of aerosols in the marine boundary layer. *J. Aerosol. Sci.* **30**, 503–532 (1999).
- Capaldo, K. P., Kasibhatla, P. & Pandis, S. N. Is aerosol production within the remote marine boundary layer sufficient to maintain observed concentrations? *J. Geophys. Res.* **104**, 3483–3500 (1999).
- Korhonen, P. et al. Ternary nucleation of H_2SO_4 , NH_3 , and H_2O in the atmosphere. *J. Geophys. Res.* **104**, 26349–26353 (1999).
- Kulmala, M., Pirjola, L. & Mäkelä, J. M. Stable sulphate clusters as a source of new atmospheric particles. *Nature* **404**, 66–69 (2000).
- Pirjola, L., O'Dowd, C. D., Brooks, I. M. & Kulmala, M. Can new particle formation occur in the clean marine boundary layer? *J. Geophys. Res.* **105**, 26531–26546 (2000).
- O'Dowd, C. D. et al. A dedicated study of new particle formation and fate in the coastal environment (PARFORCE): Overview of objectives and achievements. *J. Geophys. Res.* (in the press).
- Berresheim, H. et al. Gas-aerosol relationships of H_2SO_4 , MSA and H: Observations in the coastal marine boundary layer at Mace Head, Ireland. *J. Geophys. Res.* (in the press).
- Väkevä, M., Hämeri, K. & Aalto, P. Hygroscopic properties of nucleation mode and Aitken mode particles during and outside nucleation bursts in west coast of Ireland. *J. Geophys. Res.* (in the press).
- Mäkelä, J. M. et al. Biogenic iodine emissions and identification of end-products in coastal ultrafine particles during nucleation bursts. *J. Geophys. Res.* (in the press).
- Cocker, D. R., Flagan, R. C. & Seinfeld, J. H. State-of-the-art chamber facility for studying atmospheric aerosol chemistry. *Environ. Sci. Technol.* **35**, 2594–2601 (2001).
- Carpenter, L. J. et al. Short-lived alkyl iodides and bromides at Mace Head, Ireland: Links to biogenic sources. *J. Geophys. Res.* **104**, 1679–1689 (1999).
- Laturnus, E., Giese, B., Wiencke, C. & Adams, F. Low molecular-weight organoiodine and organobromine compounds released by polar macroalgae—the influence of abiotic factors. *Fresenius J. Anal. Chem.* **368**, 297–302 (2000).
- Giese, B., Laturnus, E., Adams, F. C. & Wiencke, C. Release of volatile iodinated $\text{C}_1\text{--C}_4$ hydrocarbons by marine macro from various climate zones. *Environ. Sci. Technol.* **33**, 2432–2439 (1999).
- Goodwin, K. D., North, W. J. & Lidstrom, M. E. Production of bromoform and dibromomethane by giant kelp: Factors affecting release and comparison to anthropogenic bromine sources. *Limnol. Oceanogr.* **42**, 1725–1734 (1997).
- Anache, S., Kamagata, Y., Kanagawa, T. & Kimuramatsu, Y. Bacteria mediate methylation of iodine in marine and terrestrial environments. *Appl. Environ. Microbiol.* **67**, 2718–2722 (2001).
- McFiggans, G. et al. A model study of iodine chemistry in the marine boundary layer. *J. Geophys. Res.* **105**, 14371–14386 (2000).
- Allan, B. J., Plane, J. M. C. & McFiggans, G. Observations of OIO in the remote marine boundary layer. *Geophys. Res. Lett.* **28**, 1945–1948 (2001).
- Dal Maso, M., Kulmala, M., Mäkelä, J. M., Aalto, P. & O'Dowd, C. D. Condensation and coagulation sinks and the formation of nucleation mode particles in coastal and boreal forest boundary layers. *J. Geophys. Res.* (in the press).
- Lushnikov, A. A. & Kulmala, M. Nucleation controlled formation and growth of disperse particles. *Phys. Rev. Lett.* **81**, 5165–5168 (1998).
- Houghton, J. T. et al. *Climate Change 2001 – The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
- Leck, C. & Bigg, E. K. Aerosol production over remote marine areas—A new route. *Geophys. Res. Lett.* **23**, 3577–3580 (1999).
- Murphy, D. M., Thomson, D. S. & Middlebrook, A. M. Bromine, iodine and chlorine in single aerosol

- particles at Cape Grim. *Geophys. Res. Lett.* **24**, 3197–3200 (1997).
25. Gähler, H. E. & Heumann, K.-G. Determination of particulate iodine in aerosols from different regions by size fractioning impactor sampling and IDMS. *Int. J. Anal. Chem.* **50**, 129–146 (1993).
 26. Wimschneider, A. & Heumann, G. Iodine: speciation in size fractionated atmospheric particles by isotope dilution mass spectrometry. *Fresenius J. Anal. Chem.* **353**, 191–196 (1995).
 27. Baker, A. R., Thompson, D., Campos, M. L. A. M., Parry, S. J. & Jickells, T. D. Iodine concentration and availability in atmospheric aerosol. *Atmos. Environ.* **34**, 4331–4336 (2000).
 28. Huang, S., Arimoto, R. & Rahn, K. A. Sources and source variations for aerosol at Mace Head, Ireland. *Atmos. Environ.* **35**, 1421–1437 (2001).
 29. Jayne, J. T. *et al.* Development of an aerosol mass spectrometer for size and composition analysis of submicron particles. *Aerosol Sci. Technol.* **33**, 49–70 (2000).
 30. Bloss, W. J., Rowley, D. M., Cox, R. A. & Jones, R. L. Kinetics and products of the IO self-reaction. *J. Phys. Chem. A* **105**, 7840–7854 (2001).

Acknowledgements

This work was funded by the European Commission, the Finnish Academy, the US Department of Energy and the National Development Plan and Environmental Protection Agency, Ireland.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.D.O'D. (e-mail: colin.odowd@cmas.demon.co.uk).

Biodiversity as a barrier to ecological invasion

Theodore A. Kennedy*, Shahid Naeem†, Katherine M. Howe‡, Johannes M. H. Knops‡, David Tilman* & Peter Reich§

* Department of Ecology, Evolution, and Behavior, University of Minnesota, 1987 Upper Buford Circle, Saint Paul, Minnesota 55108, USA

† Department of Zoology, University of Washington, 24 Kincaid Hall, Box 351800, Seattle, Washington 98195-1800, USA

‡ School of Biological Sciences, University of Nebraska, 348 Manter Hall, Lincoln, Nebraska 68588-0118, USA

§ Department of Forest Resources, University of Minnesota, 1530 Cleveland Avenue North, Saint Paul, Minnesota 55108, USA

Biological invasions are a pervasive and costly environmental problem^{1,2} that has been the focus of intense management and research activities over the past half century. Yet accurate predictions of community susceptibility to invasion remain elusive. The diversity resistance hypothesis, which argues that diverse communities are highly competitive and readily resist invasion^{3–5}, is supported by both theory⁶ and experimental studies^{7–14} conducted at small spatial scales. However, there is also convincing evidence that the relationship between the diversity of native and invading species is positive when measured at regional scales^{3,11,15,16}. Although this latter relationship may arise from extrinsic factors, such as resource heterogeneity, that covary with diversity of native and invading species at large scales, the mechanisms conferring greater invasion resistance to diverse communities at local scales remain unknown. Using neighbourhood analyses, a technique from plant competition studies^{17–19}, we show here that species diversity in small experimental grassland plots enhances invasion resistance by increasing crowding and species richness in localized plant neighbourhoods. Both the establishment (number of invaders) and success (proportion of invaders that are large) of invading plants are reduced. These results suggest that local biodiversity represents an important line of defence against the spread of invaders.

The relative contribution of extrinsic factors and resident diversity to invasion resistance varies across spatial scales^{11,12}. For example, the initial arrival of propagules of invading plants into a

formerly isolated region or field is regulated by extrinsic factors, such as human transport that breaches isolation. Similarly, if an invading plant lands on a bare patch within a field, extrinsic factors such as the frequency and intensity of disturbance, soil fertility and climate are likely to be more important than resident diversity in regulating the success of invaders. Within a field, however, in a vegetated patch, an invading plant will find itself surrounded by neighbouring plants, and it is here that biotic factors such as species composition and plant density will regulate the competitive environment the invader faces.

To test for the impacts of biodiversity on invasion resistance independent of extrinsic factors, we examined how variation in plant species richness among 147 experimental grassland plots at Cedar Creek, Minnesota, affected plant neighbourhood properties, and how these properties, in turn, affected the establishment and success of the 13 species of exotic non-native plants (primarily Eurasian, see Methods for species list) that invaded these plots. We focused on three properties of plant neighbourhoods that are widely used in studies of intra- and interspecific competition and have been shown to affect the performance of plants^{17–19}. These properties are (1) the number of species within the neighbourhood (S_N), (2) the number of plants in the neighbourhood (N) and (3) θ_{sum} (ref. 12), an index of crowding that takes into account both the distance and size of all the plants within the neighbourhood.

Studies in which invading plants are established in neighbourhoods of a fixed size and composition^{13,18} cannot test whether neighbourhood properties affect invasions in spatially heterogeneous systems like natural vegetation. We avoided this limitation by examining the neighbourhood properties of all naturally invading exotic plants, as well as the neighbourhood properties of 100 randomly placed ‘null’ positions, in each plot. The former neighbourhoods reflect the biotic environment where invaders successfully established, whereas the latter null neighbourhoods reflect the typical biotic environments found within the plot, independent of invader establishment.

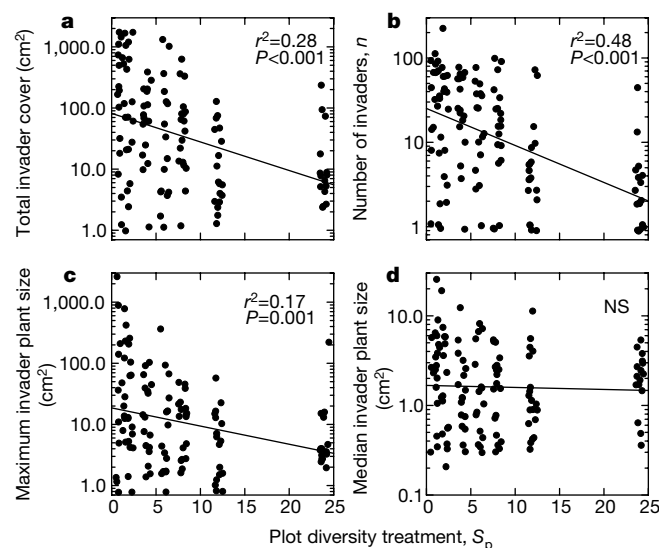


Figure 1 Plot diversity and invader performance. Effects of plot diversity treatment on total invader cover per subplot (a), the number of invading plants per subplot (b), the largest individual invader per subplot (c), and the median size of individual invaders in a subplot (d). Plot level patterns (a) are the result of a decrease in the number of invaders (b), and a reduction in the size of the largest invading plants in a subplot (c), but not a change in the median size of individual invaders (d), in more diverse plots. The least-squares regression lines are shown. Data are from 1998. Note the log scale on all y-axes; NS, not significant.

Our experimental system consisted of 147 different 3×3 m plots in which plot diversity (S_p) initially varied from 1, 2, 4, 6, 8, 12 and 24 grassland species drawn at random from a pool of 24 species²⁰. We recorded the spatial coordinates and size of approximately 53,000 individual plants, 40,000 residents and 13,000 invaders, in 40×125 cm subplots within each plot during 1997 and 1998. Three years of weeding before the invasions study, and an initial disking and removal of the topmost layer of soil when the plots were established, eliminated most of the seed bank. Weeding ceased within the subplot for the two years that we conducted our experiment. Established invaders were therefore primarily derived from newly arriving seeds from large stretches of adjacent weedy fields. Random assignment of treatments to plots within the experimental site normalized seed rain distances from adjacent fields. We analysed these invasions using high-resolution digitized maps that recorded plants as small in area as 0.1 cm^2 . We used the amount of bare ground within the subplot as a covariate in order to account for a known negative relationship between S_p and bare ground in these experimental plots²⁰, and thus test specifically for the effects of diversity on invasions. Average neighbourhood crowding and density of resident plants increased slightly across all treatment levels from 1997 to 1998, indicating that our results are not due to invader effects, such as a reduction in resident cover, as a result of not weeding during the two years of this experiment.

Plot species richness (S_p) had significant negative effects on the total area covered by invading plants, even after correcting for differences in the amount of total area covered by bare ground (analysis of covariance, ANCOVA, $F = 3.164$, $P = 0.006$; Fig. 1a). There were 91% and 96% reductions in invader cover in plots where $S_p = 24$ compared to the monocultures (that is, $S_p = 1$) in 1997 and 1998, respectively. This strong negative effect of diversity on total area covered by invaders was due to reductions in both the number of invaders (Fig. 1b) and the maximum size of individual invading plants in diverse plots (Fig. 1c), but not due to changes in the median size of individual invaders (Fig. 1d). Thus, S_p had a strong negative effect on both invader establishment (fewer invaders) and success (smaller maximum size) that was independent of bare ground. S_p also had a significant and positive effect on neighbourhood characteristics associated with crowding (Fig. 2a and b). Increasing S_p was positively associated with plant neighbourhoods that had greater S_N , greater N and increased θ_{sum} . These effects were also independent of covariance owing to bare ground (Table 1).

Invaders became established in neighbourhoods that were sig-

nificantly less crowded than 'null' neighbourhoods. The distribution of θ_{sum} for randomly placed 'null' neighbourhoods differed significantly from those neighbourhoods with invaders (Kolmogorov–Smirnov test, $P = 0.013$). Furthermore, invaders occurred in neighbourhoods with significantly lower θ_{sum} on average than 'null' neighbourhoods (t -test, $t = 2.848$, d.f. = 266, $P < 0.01$).

Increasingly dense, species-rich, and crowded neighbourhoods led to a marked decrease in the number and proportion of invading plants that attain a large size (that is, size $> 10 \text{ cm}^2$; Fig. 3a and b). Logistic regression results indicate that S_N , N and θ_{sum} are significant predictors of whether an invading plant will be large, even when the amount of bare ground in a plot and the diversity treatment of a plot are also included as predictors. (We performed a logistic regression with the categorical response variable of whether the invader is $> 10 \text{ cm}^2$. Overall $\chi^2 = 92.68$, d.f. = 4, $P < 0.001$. T -ratio for individual predictors: $S_p = 3.871$, $P < 0.001$, bare ground = 3.147, $P < 0.005$, $S_p \times$ bare ground = -4.200 , $P < 0.001$, $N = -6.610$, $P < 0.001$. N , S_N and θ_{sum} are highly correlated, and only N was included in the model to avoid problems with collinearity, but similar results are obtained for each of the neighbourhood level predictors.)

Disentangling the relative contributions of sampling, niche-complementarity, and other potential mechanisms responsible for the observed diversity effects in diversity-gradient experiments requires further experimentation²¹. However, analysis of data from the 24-species plots further supports neighbourhood properties as an important determinant of invasion resistance. In the 24-species plots, neighbourhood properties based on "null" invaders were a significant predictor of total invader cover, just as they were across the entire experimental diversity gradient, even when the plot level variables of bare ground and nitrate in the rooting zone were also included as predictors (linear regression, stepwise backward deletion, $R^2 = 0.35$, $P < 0.03$). In other words, even when plot level diversity is held constant ($S_p = 24$) and covarying extrinsic factors (bare ground and nitrate) are controlled, we still find that neighbourhood characteristics are important and significant determinants of invader success.

Collectively, these results demonstrate that greater local species richness (S_p) increases neighbourhood species richness (S_N), density (N) and crowding (θ_{sum}), and these biotic factors are clearly associated with decreases in invader establishment and success, independently of the amount of bare ground. Because all plots had the same size, history, exposure to extrinsic factors, and an unbiased seed rain from weedy fields, we suggest that the resistance

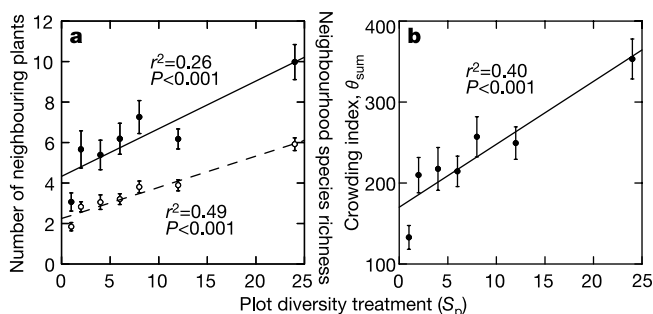


Figure 2 Neighbourhood characteristics and plot diversity treatment. Diverse plots tend to have neighbourhoods with more neighbouring plants (a, solid symbols and line), neighbourhoods that are more species rich (a, open symbols and dashed line), and neighbourhoods that are more crowded (b). Data points are mean values of 100 null points randomly placed into each of the 147 maps from 1998 that were then averaged within diversity treatments, $\pm$ one standard error.

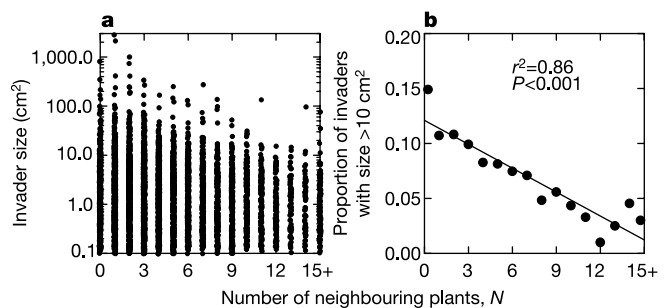


Figure 3 The influence of the number of neighbouring plants on invader performance. The upper bound of individual invader size markedly decreases as the number of neighbouring plants increases (a). Panel b depicts this decrease in upper bound in another way, by plotting the proportion of large invading plants as a function of the number of neighbouring plants (large plants were arbitrarily defined as plants $> 10 \text{ cm}^2$, a size cut-off that only 9% of all invaders met. The pattern also holds for other cut-off values and neighbourhood parameters). Data are from both 1997 and 1998.

Table 1 Effects of diversity treatment on neighbourhood properties when controlling for extent of bare ground

Dependent variables	Diversity treatment (S_P)	Bare ground
d.f.	139	139
Wilke's λ	0.420	0.361
(F , d.f., d.f., P)	(7.75, 8, 387, < 0.001)	(80.67, 3, 137, < 0.001)
	CL	UF (d.f., 139)
Neighbourhood density (N)	0.536	0.001
Neighbourhood species richness (S_N)	0.890	0.126
Crowding index (θ_{sum})	0.612	0.345
		UF (d.f., 139)
		0.000
		3.928*
		29.326***

Multivariate analysis of covariance results for the effect of plot diversity treatment and bare ground (covariate) on neighbourhood properties. Neighbourhood properties are based on the average of 100 null points placed into each plot. The highly significant effect of diversity treatment on neighbourhood properties, even after inclusion of bare ground within the subplot as a covariate, indicates that the effects of diversity on neighbourhood crowding is not simply the result of the greater plant biomass in diverse plots, but also a product of more evenly and tightly packed plants in diverse plots. Results shown are for 1998, but they are similar for other time periods. CL, canonical loadings; UF, univariate F ; d.f., degrees of freedom; θ_{sum} , crowding index (see Methods). * $P < 0.05$, *** $P < 0.001$

to invasion observed in our diverse plots is due primarily to biotic factors, namely crowded and intensely competitive plant-neighbourhoods. Interestingly, crowding and more complete use of space was also identified as the mechanism for greater invasion resistance of diverse communities of sessile marine invertebrates²².

Our study suggests that losses of biodiversity mean possible degradation of local resistance against invasion once factors that previously insulated habitats from invasion, such as geographical barriers, are breached by the changing patterns of human transport. These results also indicate that restoration and re-vegetation practitioners would benefit from establishing communities with as high a diversity of plants as is ecologically realistic and logistically feasible—diverse communities will probably require minimal maintenance and monitoring because they are generally effective at excluding undesirable invaders. □

Methods

Experimental plots

Invasers were defined as any exotic non-native plants that invaded the 147 experimental grassland plots in central Minnesota that were originally seeded with up to 24 species of perennial grassland plants²⁰. These plots are arranged in a matrix of 7 plots wide by 21 plots long, with total dimensions 24 × 75 m. Plots are a minimum of 6 m, and a maximum of 27 m, from the old fields that probably served as the seed source for invaders. The 24 species that were planted, and more details on these plots, are given elsewhere^{10,20}. Although 41 different species invaded the plots, only the 13 non-native species were included in our analyses. The non-native invaders included *Agropyron repens*, *Berteroa incana*, *Chenopodium album*, *Crepis tectorum*, *Digitaria ischaemum*, *Molluga verticillata*, *Polygonum convolvulus*, *Setaria lutescens*, *Silene antirrhina*, *Taraxacum officinale*, *Trifolium pratense*, *Trifolium repens*, and *Verbascum thapsus*.

Within the 0.5-m² experimental sub-plots where this research was conducted, between 1 and 20 species of resident plants were present. This range of species richness is comparable to the 6–22 species that were recorded in 120 different 1-m² sampling plots in a nearby undisturbed grassland²⁰, indicating that the plot diversity treatment levels roughly match those of natural plant communities. During the three years before this study, from 1994 to 1996, diversity treatments were maintained by weeding all plots at least three times during the growing season. The high frequency of weeding allowed us to remove invaders while they were still quite small and before they had the opportunity to set seed, minimizing the soil disturbance that weeding might otherwise have created. All weeding was conducted from elevated platforms, further minimizing the effects that weeding had on soil disturbance.

Neighbourhood analyses

We analysed results using a range of neighbourhood sizes (for example, 5–15-cm radius) and all yielded similar results. We present results using neighbourhoods with a 10-cm radius because this size is comparable to published values for neighbourhood analyses of plants that are similar in size to many of the invaders in our experiment¹⁸. θ is the angle subtended at an invader by another plant, and θ_{sum} is the sum of all angles for a given plant¹². To ensure that all invaders had neighbourhoods that were completely mapped, and to minimize any possible edge effects on the distribution of invaders within the subplot, only those invaders that were at least 10 cm from all edges of the subplot were included in neighbourhood analyses. This approach allowed us to monitor the success of a large number of naturally occurring individual invaders and total invader cover within a plot, and collect information on the properties of plant neighbourhoods across a range of species richness. A uniform random distribution was used to determine the placement of 'null' points. The first 100 points selected for each plot were used, regardless of whether they fell near an actual invader or not.

Statistical methods

We used multivariate analyses of variance to test for possible multivariate effects not

detectable with univariate methods. We used logistic regression to determine if neighbourhood characteristics and/or plot level characteristics are significant predictors of the performance of individual invaders. All patterns were comparable for both 1997 and 1998. We present data from only 1998 except for Fig. 3, which includes data from both years.

Received 30 January; accepted 20 March 2002; doi:10.1038/nature00776.

1. Mooney, H. A. & Drake, J. A. *Ecology of Biological Invasions of North America and Hawaii* (Springer, New York, 1986).
2. Vitousek, P. M., D'Antonio, C. M., Loope, L. L. & Westbrooks, R. Biological invasions as global environmental change. *Am. Sci.* **84**, 468–478 (1996).
3. Levine, J. & D'Antonio, C. M. Elton revisited: a review of evidence linking diversity and invasibility. *Oikos* **87**, 15–26 (1999).
4. Crawley, M. J. in *Colonization, Succession and Stability* (eds Crawley, M. J., Edwards, P. J. & Gray, A. J.) Blackwell, Oxford, (1987), 429–453.
5. Elton, C. S. *The Ecology of Invasions by Animals and Plants* (Methuen, London, 1958).
6. Case, T. J. Invasion resistance arises in strongly interacting species-rich model competition communities. *Proc. Natl Acad. Sci. USA* **87**, 9610–9614 (1990).
7. Tilman, D. The ecological consequences of changes in biodiversity: A search for general principles. *Ecology* **80**, 1455–1474 (1999).
8. McGrady-Steed, J., Harris, P. & Morin, P. Biodiversity regulates ecosystem reliability. *Nature* **390**, 162–165 (1997).
9. Tilman, D. Community invasibility, recruitment limitation, and grassland biodiversity. *Ecology* **78**, 81–92 (1997).
10. Knops, J. M. H. et al. Effects of plant species richness on invasion dynamics, disease outbreaks, insect abundances and diversity. *Ecol. Lett.* **2**, 286–293 (1999).
11. Levine, J. Plant diversity and biological invasions: Relating local process to community pattern. *Science* **288**, 852–854 (2000).
12. Naeem, S. et al. Plant diversity increases resistance to invasion in the absence of covarying extrinsic factors. *Oikos* **91**, 97–108 (2000).
13. Symstad, A. J. A test of the effects of functional group richness and composition on grassland invasibility. *Ecology* **81**, 99–109 (2000).
14. Dukes, J. S. Biodiversity and invasibility in grassland microcosms. *Oecologia* **126**, 563–568 (2001).
15. Planty-Tabacchi, A. M., Tabacchi, E., Naiman, R. J., Deferrari, C. & Decamps, H. Invasibility of species-rich communities in riparian zones. *Conserv. Biol.* **10**, 598–607 (1996).
16. Stohlgren, T. J. et al. Exotic plant species invade hot spots of native plant diversity. *Ecol. Monogr.* **69**, 25–46 (1999).
17. Mack, R. N. & Harper, J. L. Interference in dune annuals: spatial pattern and neighbourhood effects. *J. Ecol.* **65**, 345–364 (1977).
18. Pacala, S. W. & Silander, J. A. Jr Neighbourhood interference among velvet leaf, *Abutilon theophrasti*, and pigweed, *Amaranthus retroflexus*. *Oikos* **48**, 217–224 (1987).
19. Rejmanek, M. & Messina, J. J. *Multiresource Management of Ponderosa Pine Forests* General Technical Report RM-185 (eds Tecle, A., Covington, W. W. & Haure, R. H.) 97–102 (United States Department of Agriculture, Ft. Collins, 1989).
20. Tilman, D., Wedin, D. & Knops, J. M. H. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
21. Wardle, D. A. Experimental demonstration that plant diversity reduces invisibility—evidence of a biological mechanism or a consequence of sampling effect? *Oikos* **95**, 161–170 (2001).
22. Stachowicz, J. J., Whitlatch, R. B. & Osman, R. W. Species diversity and invasion resistance in a marine ecosystem. *Science* **286**, 1577–1579 (1999).

Acknowledgements

We thank J. Craine for suggestions with analyses, S. Hobbie, J. Fargione and J. Craine for comments on the manuscript, and T. Mielke and the undergraduate interns at Cedar Creek who made this work possible. This work was supported by the Andrew Mellon Foundation, the McKnight Foundation, the University of Minnesota, NASA and the NSF.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.A.K. (e-mail: kenn0148@tc.umn.edu).

Plant biomarkers in aerosols record isotopic discrimination of terrestrial photosynthesis

Maureen H. Conte & John C. Weber

Woods Hole Oceanographic Institution, Department of Marine Chemistry and Geochemistry, Woods Hole, Massachusetts 02543, USA

Carbon uptake by the oceans and by the terrestrial biosphere can be partitioned using changes in the $^{12}\text{C}/^{13}\text{C}$ isotopic ratio ($\delta^{13}\text{C}$) of atmospheric carbon dioxide^{1–4}, because terrestrial photosynthesis strongly discriminates against $^{13}\text{CO}_2$, whereas ocean uptake does not. This approach depends on accurate estimates of the carbon isotopic discrimination of terrestrial photosynthesis (Δ ; ref. 5) at large regional scales⁶, yet terrestrial ecosystem heterogeneity⁷ makes such estimates problematic. Here we show that ablated plant wax compounds in continental air masses can be used to estimate Δ over large spatial scales and at less than monthly temporal resolution. We measured plant waxes in continental air masses advected to Bermuda, which are mainly of North American origin, and used the wax isotopic composition to estimate Δ simply. Our estimates indicate a large (5–6‰) seasonal variation in Δ of the temperate North American biosphere, with maximum discrimination occurring in late spring, coincident with the onset of production. We suggest that the observed seasonality arises from several factors, including seasonal shifts in the proportions of production by C_3 and C_4 plants, and environmentally controlled adjustments in the photosynthetic discrimination of C_3 -plant-dominated ecosystems.

Although generally treated as a constant in carbon cycle models^{1–4}, Δ in terrestrial ecosystems varies both spatially and temporally^{7–12}. Coupled biosphere–climate models^{6,13} and analyses of seasonal variations in atmospheric CO_2 concentration (ref. 14) and $\delta^{13}\text{CO}_2$ have been used to estimate the magnitude and variability of Δ on large regional scales, but these approaches incorporate substantial uncertainties and their results are largely unverified. Here we introduce a molecular ‘biomarker’-based technique that uses the ubiquitous presence of plant leaf waxes in well-mixed continental air masses to isolate and quantify a discrete molecular signal of terrestrial photosynthetic discrimination. All higher plants produce epicuticular waxes, which cover leaf surfaces¹⁵. Wax biosynthesis is an actively regulated process with substantial wax turnover as leaves age^{15,16}. Wind and dust periodically ablate waxes from vegetation^{17,18}. Ablated waxes accumulate in air as micrometre-sized¹⁹ particles with a molecular composition broadly similar to that of their source vegetation^{17,20,21}. Wax aerosols are removed from air mainly by rain washout²². Hence, air masses moving across the land surface can accumulate and integrate waxes from vegetation over large continental areas.

We measured leaf waxes in continental air masses advected over the western North Atlantic ocean by large-scale weather systems. Our measurement site was the Atmosphere–Ocean Chemistry Experiment (AEROCE)²³ tower located on the southwestern coast of Bermuda. The region is mainly influenced by North American air masses advected by westerly winds, although North African and occasionally European air masses advected by easterly winds also occur, especially in summer and autumn²³. Transport times are typically a few days for North American air masses and about a week for North African and European air masses^{23,24}. We collected bulk aerosol samples from November 1995 to August 1998. Samples were collected on 1- μm combusted (380 °C) Whatman A/E glass-fibre filters using a high-volume air sampler that pumped when the wind

was from the ocean sector (180–330°) and between 5 and 40 knots²⁴. Average sampling duration was 14 days. Air mass trajectories (calculated twice per day) indicated that each sample was generally an admixture of aerosols from multiple air masses with widely varying continental trajectories. We analysed the molecular composition of the leaf wax and the isotopic composition of individual leaf wax compounds by gas chromatography (GC) and gas chromatography–isotope ratio mass spectrometry (GC–IRMS) using previously described methods²⁴.

The mean atmospheric concentrations of the C_{20-34} *n*-alkanols and *n*-alkanoic acids, the major leaf wax compound classes, are shown in Fig. 1a. Concentration variations reflect the episodic nature of long-range transport rather than any seasonality of wax production. Figure 1b shows the $\delta^{13}\text{C}$ values of individual compounds and Fig. 1c shows the concentration-weighted average $\delta^{13}\text{C}$ values ($\delta^{13}\text{C}_{\text{avg}}$) of the even C_{24-32} compounds for each class. The magnitude and range of leaf wax $\delta^{13}\text{C}$ values observed in the

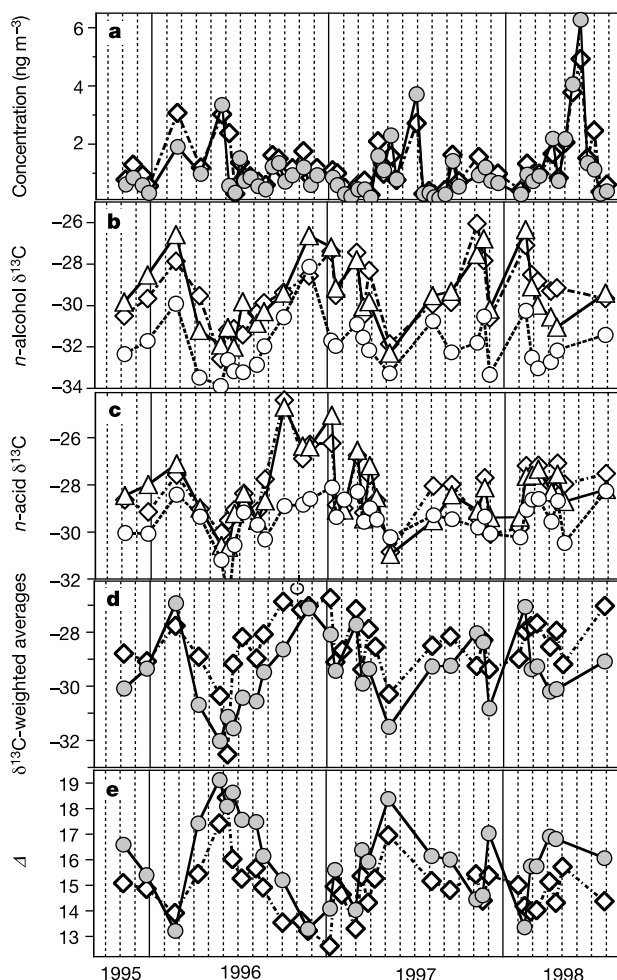


Figure 1 Time series of concentrations and carbon isotopic ratios ($\delta^{13}\text{C}$) of leaf wax compound classes in Bermuda aerosols. Solid vertical lines mark the beginning of each year and dashed vertical lines the beginning of each month. **a**, Concentrations of total C_{20-34} *n*-alkanols (circles, solid line) and C_{20-34} *n*-acids (diamonds, dashed line). Data are plotted at the midpoint of each sampling interval. **b**, **c**, $\delta^{13}\text{C}$ values for individual C_{26} (circles, dashed line), C_{30} (diamonds, dotted and dashed line) *n*-alcohol and *n*-acid compounds. Data are reported relative to the Pee Dee Belemnite standard. **d**, Concentration-weighted average values of $\delta^{13}\text{C}$ for even C_{24-32} *n*-alkanols and *n*-acids. **e**, Derived estimates of Δ as described in the text. Symbols for **d** and **e** are as in **a**.

aerosols is within that reported for C_3 plants^{21,25,26}, which is consistent with the dominance of C_3 plant ecosystems in North America, the primary source region²⁴. (Samples filtering North African air masses containing large quantities of Saharan dust also exhibited isotopic compositions indicative of C_3 plant dominance, which suggests that the waxes in African air masses advected to the western North Atlantic mainly originate in C_3 forests south of the dust source areas²⁴.) Although we observed molecular variations in our samples indicative of differences in source vegetation²⁴, isotopic variations were not correlated with either molecular composition or air mass trajectories. Rather, wax isotopic variation was dominated by a large (5–6‰) seasonal cycle that essentially overwhelmed any more subtle isotopic differences that could be attributed to source differences. We did observe, however, that n -alkanoic acids were about 1‰ more enriched in ^{13}C than n -alkanols. This might reflect a smaller biosynthetic isotopic fractionation (ϵ_f) of n -alkanoic acids, although existing data^{21,25} and biosynthetic considerations¹⁵ suggest that ϵ_f is similar in both wax classes. Alternatively, this could reflect different source strengths among the vegetation types that contribute these wax classes. For example, angiosperm waxes contain both n -alkanoic acids and n -alkanols, whereas gymnosperm waxes are depleted in n -alkanols¹⁵ and so contribute mainly n -alkanoic acids.

We converted the $\delta^{13}C_{avg}$ of the n -alkanols and n -alkanoic acids to Δ using the formula $\Delta = (\delta a - \delta x)/(1 + \delta x)^5$, where δa and δx are the isotopic composition (in absolute units) of the source CO_2 and bulk leaf carbon, respectively. For δa , we used seasonal measurements of $\delta^{13}CO_2$ in boundary-layer air at a North Carolina site¹⁴. For δx , we used the formula $\delta x = \delta_{wax} + \epsilon_f$, where δ_{wax} is the measured $\delta^{13}C_{avg}$ of the n -alkanols or n -alkanoic acids and ϵ_f is the mean biosynthetic isotopic fractionation between the leaf wax compound class and bulk leaf carbon. Few compound-specific isotopic measurements of plant waxes exist. Studies of dominant North American ecosystems are underway to constrain the magnitude of ϵ_f ; in the meantime, we have used a constant estimate of 6‰ (± 0.7 ‰) for both n -alkanols and n -acids. This estimate is based on our multi-year data on wax isotopic composition in biomass in a northern C_3/C_4 mixed grass prairie²¹ and previously

published isotopic data^{25,26} on wax ϵ_f in other C_3 species. Existing data indicate that some C_4 plants have a larger ϵ_f (~ 8 – 10 ‰)^{21,25,26}, which might result in small overestimates of Δ in samples having significant inputs from C_4 plants.

The time series of Δ estimated from the n -alkanol and n -alkanoic acid isotopic data sets are shown in Fig. 1e. Estimates strongly covary for the two classes, although the estimate for n -alkanoic acids is about 1‰ lower for reasons (discussed above) that are not clear. The magnitude of Δ that we estimate ranges from a minimum of about 13‰ in winter to a maximum of 18–19‰ in late spring. This range is similar to that previously reported for C_3 -plant-dominated ecosystems^{6–13}, and provides validation of the wax aerosol approach and our estimated magnitude of ϵ_f . The concentration-weighted mean Δ for the entire time series is 16.2‰ using the n -alkanols and 15.1‰ using n -alkanoic acids, which is within the range of model estimates¹³ for annually averaged Δ in North America.

The seasonal cycle of Δ estimated using the n -alkanol data is shown in Fig. 2. Estimates for samples collected in the same month but in different years are markedly similar, despite evidence from air mass trajectories that these samples collected waxes from air masses that typically had very different continental trajectories²⁴. This result strongly indicates that Bermuda aerosols are recording a signal of Δ that is representative of the large regional to subcontinental average, and implies that the atmospheric residence time of ablated waxes is sufficient to integrate inputs over large expanses. This conclusion is consistent with evidence^{19,20,24,27} for the ubiquitous presence of plant waxes in air, even in remote ocean settings.

We compared monthly averages of Δ estimated from the n -alkanol and n -alkanoic acid aerosol data and monthly estimates of net primary production (NPP) for temperate North America and the southeastern USA (Fig. 3), which were computed using the Carnegie Ames Stanford Approach (CASA) model²⁸. The seasonal cycles of Δ and NPP strongly covary, indicating a tight link between

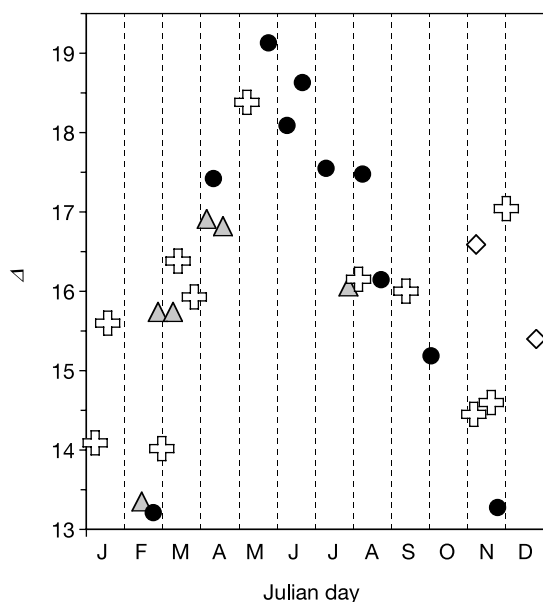


Figure 2 The seasonal cycle of Δ estimated from plant wax n -alkanols in Bermuda aerosols. Four consecutive years are overlain on the same time axis: 1995 (diamonds), 1996 (circles), 1997 (crosses) and 1998 (triangles).

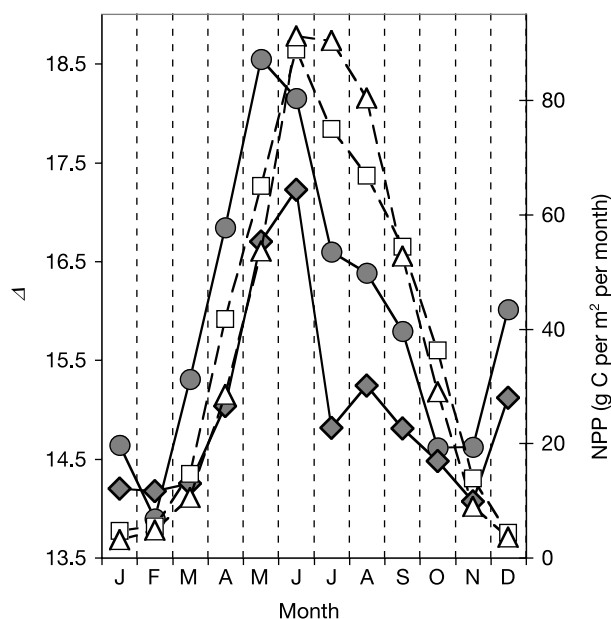


Figure 3 Comparison of monthly averages of Δ estimated from plant wax n -alkanols (circles, solid line) and n -acids (diamonds, solid line) in Bermuda aerosols, and modelled estimates of net primary production (NPP) using the CASA²⁸ terrestrial biosphere model. Monthly NPP estimates are shown for temperate North America (25–50° N, 50–150° W; triangles, dashed line) and for the southeastern USA (25–40° N, 70–90° W; squares, dashed line).

Δ and terrestrial productivity cycles on continental scales. In addition, the covariance and phasing of the seasonal cycle of Δ , as recorded by the Bermuda aerosols, and North American NPP provides strong evidence that the temporal lag between wax biosynthesis and transport to our measurement site is on the order of weeks rather than months. The seasonal cycle of the estimate for *n*-alkanoic acids lags that of the *n*-alkanols by about 1 month. The reason is not clear but may reflect differences in source inputs. The rise in Δ in early spring closely parallels the rise in estimated NPP and suggests that photosynthetic discrimination is greatest at the onset of the growing season. The decline in estimates of Δ from both *n*-alkanols and *n*-alkanoic acids closely parallels the decline in NPP, especially in the southern USA.

We calculated an annually integrated estimate of Δ using the monthly NPP estimates (Fig. 3) to flux-weight our data. Our calculation assumes that the temporal lag between wax biosynthesis and collection is within our sampling resolution. Our NPP-weighted annual averages of Δ are 16.7‰ and 14.6‰ using the data for *n*-alkanols and *n*-alkanoic acids, respectively (about 0.3‰ of the difference in estimates arises from the anomalously low *n*-acid $\delta^{13}\text{C}$ in July). The *n*-alkanol estimate is nearly identical to previous estimates¹⁴ of Δ (15.7–17.1‰, mean 16.6‰, which we calculated from ref. 14's reported data for $\delta_{\text{bio}}^{13}\text{C}$ computed from seasonal variations in CO_2 concentration and $\delta^{13}\text{C}_{\text{CO}_2}$ (corrected for fossil fuel contamination) at several North America sites and Bermuda. Our estimate also agrees with model estimates¹⁵ of annual North American ecosystem Δ computed using gridded vegetation and climate data. This general agreement provides additional support that the isotopic signal of leaf waxes in the Bermuda aerosols reflects a large-scale North American biosphere signal of Δ with minimal isotopic bias.

The 5–6‰ seasonal amplitude in Δ indicated here has been previously predicted for the temperate latitudes⁶, but has not previously been confirmed by direct measurement. It is likely that the seasonality arises from a combination of factors. First, plant studies^{7–9,11} and models^{5,6,13} suggest that isotopic discrimination of C_3 plants is highest during the optimal growth conditions of early spring, and declines later in the growing season as stomatal conductance is reduced owing to moisture deficits and leaf senescence. Seasonal declines in the overall discrimination within C_3 ecosystems are also indicated by studies¹⁰ which find that plants with highest Δ are more active in moisture-replete conditions typical of spring, whereas plants with lower Δ perform better during moisture-limiting conditions typical of late summer. Second, production by C_4 plants is closely tied to temperature and becomes an increasingly larger fraction of total ecosystem production as the growing season progresses^{29,30}. The large spatially integrated signal at Bermuda also incorporates seasonal changes in source strengths of ecosystems at different latitudes. Low-latitude ecosystems, which generally contain more C_4 plant species²⁹ and are modelled to have a lower overall Δ ^{6,13}, contribute a proportionately greater fraction of the signal early and late in the year when higher-latitude ecosystems are dormant. Further studies are needed to quantitatively assess the contributions of these factors to large-scale seasonality in Δ .

The leaf wax–aerosol method introduced here exploits the specificity of organic molecules and the integrating properties of atmospheric mixing to overcome the limitations inherent in scaling heterogeneous terrestrial ecosystem processes to large spatial scales. Method approximations are straightforward and easily refined as new information becomes available. The simple sampling technique makes the method well suited for sustained observations. Application of the leaf wax–aerosol method at key sites in conjunction with measurements of CO_2 concentration and $\delta^{13}\text{C}_{\text{CO}_2}$ can provide data on Δ that are needed to constrain model estimates of carbon sequestration, as well as important insights about the large-scale functioning of the terrestrial biosphere. □

Received 3 August 2001; accepted 8 March 2002; doi:10.1038/nature00777.

1. Tans, P. P., Berry, J. A. & Keeling, R. F. Oceanic $^{13}\text{C}/^{12}\text{C}$ observations: a new window on oceanic CO_2 uptake. *Global Biogeochem. Cycles* **7**, 353–368 (1993).
2. Ciais, P. *et al.* Partitioning of ocean and land uptake of CO_2 as inferred by $\delta^{13}\text{C}$ measurements from the NOAA Climate Modeling and Diagnostics Laboratory global air sampling network. *J. Geophys. Res.* **100**, 5051–5070 (1995).
3. Francey, R. J. *et al.* Changes in oceanic and terrestrial carbon uptake since 1982. *Nature* **373**, 326–330 (1995).
4. Rayner, P. J., Enting, I. G., Francey, R. J. & Langefelds, R. Reconstructing the recent carbon cycle from atmospheric CO_2 , $\delta^{13}\text{C}$ and O_2/N_2 observations. *Tellus* **51B**, 213–232 (1999).
5. Farquhar, G. D. & Lloyd, J. in *Stable Isotopes and Plant Carbon–Water Relations* (eds Ehleringer, J. R., Hall, A. H. & Farquhar, G. D.) 47–70 (Academic, San Diego, 1993).
6. Fung, I. *et al.* Carbon 13 exchanges between the atmosphere and biosphere. *Global Biogeochem. Cycles* **11**, 507–533 (1997).
7. Ehleringer, J. R., Hall, A. E. & Farquhar, G. D. (eds) *Stable Isotopes and Plant Carbon–Water Relations* (Academic, San Diego, 1993).
8. Flanagan, L. B. & Ehleringer, J. R. Ecosystem–atmosphere CO_2 exchange: interpreting signals of change using stable isotopic ratios. *Trends Ecol. Evol.* **13**, 10–14 (1998).
9. Marshall, J. D. & Zhang, J. Carbon isotope discrimination and water use efficiency in native plants of the north central Rockies. *Ecology* **75**, 1887–1895 (1994).
10. Ehleringer, J. R. Variation in leaf carbon isotope discrimination in *Encelia farinosa*: implications for growth, competition, and drought survival. *Oecologia* **95**, 340–346 (1993).
11. Smedley, M. P. *et al.* Seasonal carbon isotopic discrimination in a grassland community. *Oecologia* **85**, 314–320 (1991).
12. Körner, Ch., Farquhar, G. D. & Wang, S. C. Carbon isotopic discrimination by plants follows latitudinal and altitudinal trends. *Oecologia* **88**, 30–40 (1991).
13. Lloyd, J. & Farquhar, G. D. $\delta^{13}\text{C}$ discrimination during CO_2 assimilation by the terrestrial biosphere. *Oecologia* **99**, 201–215 (1994).
14. Bakwin, P. S., Tans, P. P., White, J. W. C. & Andes, R. J. Determination of the isotopic $^{13}\text{C}/^{12}\text{C}$ discrimination by terrestrial biology from a global network of observations. *Global Biogeochem. Cycles* **12**, 555–562 (1998).
15. Gulz, P.-G. Epicuticular leaf waxes in the evolution of the plant kingdom. *J. Plant Physiol.* **143**, 453–464 (1994).
16. Lockheart, M. J., Van Bergen, P. F. & Evershed, R. P. Variations in the stable isotopic compositions of individual lipids from the leaves of higher angiosperms: implications for the study of higher plant-derived sedimentary organic matter. *Org. Geochem.* **26**, 137–153 (1997).
17. Rogge, W. F., Hildemann, L. M., Mazurek, M. A., Cass, G. R. & Simoneit, B. R. T. Sources of fine organic aerosol 4, Particles shed from leaf surface of higher plants. *Environ. Sci. Technol.* **28**, 1375–1388 (1993).
18. Hadley, J. L. & Smith, W. K. Wind erosion of leaf surface wax in alpine timberline conifers. *Arctic Alpine Res.* **21**, 392–398 (1989).
19. Peltzer, E. T. & Gagosian, R. B. in *Chemical Oceanography* (eds Riley, J. P., Chester, R. & Duce, R. A.) Vol. 10 282–338 (Academic, San Diego, 1989).
20. Simoneit, B. R. T. Organic matter of the troposphere—V: Application of molecular marker analysis to biogenic emissions into the troposphere for source reconciliations. *J. Atmos. Chem.* **8**, 251–275 (1989).
21. Conte, M. H., Weber, J. C., Flanagan, L. P. & Carlson, P. Groundtruthing the CSIA-aerosol technique for estimation of carbon isotopic discrimination of terrestrial photosynthesis. *EOS Trans. AGU* **82**(20), S77 (2001).
22. Zafriou, O. C., Gagosian, R. B., Peltzer, E. T. & Almond, J. B. Air-to-sea fluxes of lipids at Enewetak Atoll. *J. Geophys. Res.* **90**, 2409–2423 (1985).
23. Moody, J. L., Oltman, S. J., Levy, H. & Merrill, J. T. Transport climatology of tropospheric ozone: Bermuda, 1988–1991. *J. Geophys. Res.* **100**, 7179–7194 (1995).
24. Conte, M. H. & Weber, J. C. Long range atmospheric transport of terrestrial organic biomarkers to the western North Atlantic. *Org. Geochem. Special issue Advances in Organic Geochemistry 2001* (in the press).
25. Collister, J. W., Rieley, G., Stern, B., Eglinton, G. & Fry, B. Compound specific isotopic ^{13}C analyses of leaf lipids from plants with differing carbon dioxide metabolism. *Org. Geochem.* **21**, 619–628 (1994).
26. Ballentine, D. C., Macko, S. A. & Turekian, V. C. Variability of stable carbon isotopic compositions in individual fatty acids from combustion of C_4 and C_3 plants: implications for biomass burning. *Org. Geochem.* **25**, 97–104 (1998).
27. Kawamura, K. in *Biogeochemical Processes and Ocean Flux in the Western Pacific* (eds Sakai, H. & Nozaki, Y.) 31–51 (Terra Science, Tokyo, 1995).
28. Randerson, J. T., Thompson, M. V., Conway, T. J., Fung, I. Y. & Field, C. The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide. *Global Biogeochem. Cycles* **11**, 535–560 (1997).
29. Ehleringer, J. R., Cerling, T. E. & Helliker, B. R. C_4 photosynthesis, atmospheric CO_2 , and climate. *Oecologia* **112**, 285–299 (1997).
30. Still, C. J., Berry, J. A., Carbo, M. R., Fu, W. & Verma, S. B. Seasonal changes in photosynthesis and respiration from a mixed C_3 – C_4 tallgrass prairie. *EOS Trans. AGU* **80**, F93 (1999).

Acknowledgements

We thank AEROCE investigators J. Prospero, H. Maring, J. Merrill and M. McKay for assistance with sampling and sharing unpublished data; C. Johnson and L. Houghton for mass spectrometric analyses; and J. Randerson for sharing CASA model data. L. Flanagan, J. Farrington, J. Hayes, J. Fessenden and J. Randerson provided comments and suggestions that greatly improved the paper.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.H.C. (e-mail: mconte@whoi.edu).

Host-induced epidemic spread of the cholera bacterium

D. Scott Merrell^{*†}, Susan M. Butler^{*}, Firdausi Qadri[‡], Nadia A. Dolganov[§], Ahsfaul Alam[‡], Mitchell B. Cohen^{||}, Stephen B. Calderwood[¶], Gary K. Schoolnik[§] & Andrew Camilli^{*}

^{*} Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA

[‡] International Centre for Diarrhoeal Disease Research, Mohakhali, Dhaka 1212, Bangladesh

[§] Stanford Medical School, Beckman Center, Room 241, Stanford, California 94305, USA

^{||} Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA

[¶] Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts 02114, and Harvard Medical School, Boston, Massachusetts 02115, USA

The factors that enhance the transmission of pathogens during epidemic spread are ill defined. Water-borne spread of the diarrhoeal disease cholera occurs rapidly in nature, whereas infection of human volunteers with bacteria grown *in vitro* is difficult in the absence of stomach acid buffering¹. It is unclear, however, whether stomach acidity is a principal factor contributing to epidemic spread². Here we report that characterization of *Vibrio cholerae* from human stools supports a model whereby human colonization creates a hyperinfectious bacterial state that is maintained after dissemination and that may contribute to epidemic spread of cholera. Transcriptional profiling of *V. cholerae* from stool samples revealed a unique physiological and behavioural state characterized by high expression levels of genes required for nutrient acquisition and motility, and low expression levels of genes required for bacterial chemotaxis.

Frequent cholera outbreaks occur in Dhaka, Bangladesh, providing the potential to study the bacterium within its natural setting. We collected stool samples from patients at the International Centre for Diarrhoeal Disease Research (ICDDR), Bangladesh, and performed serological analysis to identify stools that were positive for *V. cholerae* O1 Inaba El Tor. These were immediately used in murine infection studies. A strain grown *in vitro* (DSM-V984) was mixed with stool *V. cholerae* and the mixture was used to inoculate 3–5-day-old Swiss Webster mice by gavage. DSM-V984 is an O1 Inaba El Tor strain that was isolated the previous year at the ICDDR, and is marked by deletion of the *lacZ* gene to allow for enumeration of stool (*LacZ*⁺) and *in vitro*-grown (DSM-V984; *LacZ*[−]) bacteria. Mice were euthanized after 20–24 h, and bacteria were recovered by homogenization of the small intestine and plating on medium supplemented with a chromogenic substrate for *LacZ*. The output ratios of stool-sample *V. cholerae* to DSM-V984 were corrected for variations in the input ratios and represent the competitive indices (CI) of the stool *V. cholerae*. A CI above 1 indicates enhanced infectivity, whereas an index below 1 indicates decreased infectivity. As shown in Fig. 1a, *V. cholerae* shed from the human gastrointestinal tract (human-shed) showed greatly enhanced infectivity, outcompeting the *in vitro*-grown strain by as much as 700-fold. A competitive advantage was consistently observed with samples collected from numerous patients on different days. The competitive advantage was entirely lost, however, when *V. cholerae* was colony purified from a stool sample and cultured *in vitro* in broth for 18 h before infection (CI geometric means of 1.1, 1.1 and 1.5 for three independent stool isolates and 8–10 mice per competition), indicating that passage through the human host transiently increased the infection potential of the stool-derived bacteria.

To have a significant role in the epidemic spread of cholera, the competitive advantage of stool-derived bacteria would have to persist after dissemination back into the environmental reservoir. We therefore tested whether human-shed *V. cholerae* maintained the hyperinfectious phenotype after incubation in local pond water. Freshly shed *V. cholerae* positive stool samples were diluted into pond water that was free of *V. cholerae* contamination. After incubation at room temperature (29 °C) for 5 h, the diluted samples were mixed with the *in vitro*-grown competitor strain, and this mixture was used to infect mice as described above. As shown in Fig. 1b, the hyperinfectious state of the human-shed bacteria remained intact. We propose that human passage of *V. cholerae* enhances subsequent water-borne spread of the cholera bacterium by effectively lowering the infectious dose in secondary individuals.

As the increased infectivity of stool-derived *V. cholerae* was lost after growth *in vitro*, the phenotype is transiently expressed. Therefore, to characterize the hypervirulent state we conducted transcriptional profiling of human-shed *V. cholerae* using a spotted DNA microarray³ containing representations of approximately 87% of the identified open reading frames (ORFs) of the Bangladeshi O1 El Tor reference strain N16961 (ref. 4). O1 Inaba El Tor positive stool samples were obtained from three patients in the ICDDR, filtered to remove particulate matter, then immediately frozen for later preparation of total RNA. These samples showed the canonical 'rice water' appearance that characterizes cholera stools, contained approximately 100 million *V. cholerae* per millilitre, and showed relatively few contaminating bacteria as assessed by dark-field microscopy and plating⁵. Stool RNA preparations were analysed by agarose gel electrophoresis to ensure RNA integrity and the absence of contaminating eukaryotic RNA. DSM-V999 was isolated from one of the three stool samples, and total RNA was isolated after *in vitro* growth to stationary phase (growth conditions identical to those used for the growth of DSM-V984 in the competition assays described above). One microgram of RNA from each of the samples was used for complementary DNA synthesis, labelled with Cy5 and hybridized to the microarray in the presence of a Cy3-labelled common reference (exponentially growing O1 Inaba El Tor strain 92A1552). Each sample was hybridized in quadruplicate and the relative fluorescent intensities determined by scanning with an Axon scanner. Data were quantified, normalized and corrected to yield the relative transcript abundance of each gene as an intensity ratio with respect to that of the reference signal. These intensity ratios were then used to identify statistically significant differences

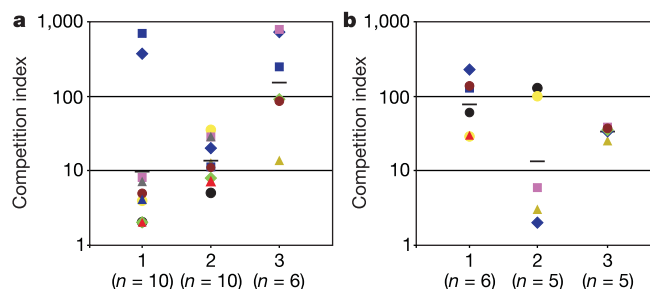


Figure 1 Human-shed *V. cholerae* are hyperinfectious in competition assays in infant mice. **a, b**, Freshly isolated stool samples from six patients (representing experiments 1, 2 and 3 on the x axes in **a** and **b**) were either mixed directly with *in vitro*-grown DSM-V984 (**a**), or incubated in local pond water for 5 h before mixing (**b**), and then used for intragastric infection of animals. Data points represent the output ratio of stool-derived *V. cholerae* to *in vitro*-grown competitor strain after correction for deviations in input ratios. Each data point represents the output from a single animal and symbols and colours are used to facilitate visual discrimination of data values. The geometric mean is indicated by a horizontal bar. *n*, total number of animals per experiment.

[†] Present address: Department of Microbiology and Immunology, Stanford School of Medicine, Stanford, California 94305, USA.

in gene expression using the Statistical Analysis for Microarrays (SAM) program⁶. A two-class SAM analysis was conducted using the strain grown *in vitro* as class I, and each individual stool sample as class II. Genes with statistically significant changes in the level of expression—at least a twofold change—in each patient sample were chosen, and the derived data from individual stool samples were collapsed to identify genes that were differentially regulated in all three samples. According to these criteria, 237 genes were differentially regulated: 44 genes were induced and 193 genes were repressed in human-shed *V. cholerae* (Fig. 2 and Supplementary Information).

The transcriptomes (the complete collection of transcribed RNAs) of the stool-derived *V. cholerae* samples were similar to that of strain DSM-V999 cultured in rich medium, as 3,120 out of 3,357 ORFs examined were expressed at roughly equivalent levels. Notable differences included increased expression of an assortment of genes required for biosynthesis of amino acids, iron uptake systems, ribosomal proteins, and formation of a periplasmic nitrate reductase complex that may allow for respiration under low oxygen tension. This transcriptome is consistent with bacterial growth in an oxygen- and iron-limited host compartment—the conditions encountered in rice-water stools⁷. Combining these observations, we propose that *V. cholerae* departs from a rich growth environment within the small intestine and moves into a nutrient poor luminal fluid that is rapidly purged. The physiological growth state of the stool bacteria and the capability for rapid growth could contribute to the observed hyperinfectious phenotype, as rapidly growing *V. cholerae* out-compete stationary phase cells (A.C., unpublished data, and J. J. Mekalanos, personal communication). However, hierarchical clustering of stationary phase, exponential phase and human-shed *V. cholerae* arrays revealed that *V. cholerae* isolated from patients cluster no closer to the exponential phase than to the

stationary phase arrays, suggesting that on a genome-wide scale, the transcriptome bears traits of both of the growth phases (data not shown).

None of the genes that showed differential regulation were members of the ToxR/TcpP/ToxT virulence gene regulon^{8–11}. Genes within this regulon include those for cholera toxin and factors coded for on the *Vibrio* pathogenicity island, such as the toxin co-regulated pilus. These factors are necessary for successful infection of humans and mice¹². These results suggest that, before being shed, *V. cholerae* turns off expression of these particular virulence genes as part of a programme for dissemination to the environment and possible transmission to new hosts. These findings also indicate that the increased expression of this virulence regulon is not required for the increased infectivity exhibited by stool bacteria.

Among the genes differentially expressed in the stool samples are VC0028, VC0941, VC0869, VC0051, VC0647, VC0468, VC2350 and VCA0583, all of which were recently identified in a genetic screen for factors required for infection of infant mice¹³. VC0647 (*pnp*) and VC0468 (*gshB*) were additionally shown to have a role in the ability of *V. cholerae* to mount an acid tolerance response (ATR). *In vitro* induction of the ATR enhances the infectivity of *V. cholerae*¹⁴; however, our microarray analysis revealed that other genes known to be induced during the ATR were not differentially regulated in stool samples, suggesting that if ATR has an involvement in the increased infectivity of human-shed *V. cholerae*, it is through pathways yet to be elucidated.

The most highly induced gene that we identified is one that was not annotated on completion of the *V. cholerae* genome sequence. As our *V. cholerae* microarray was constructed before annotation of the published sequence¹⁵, it contains sequences of some ORFs that

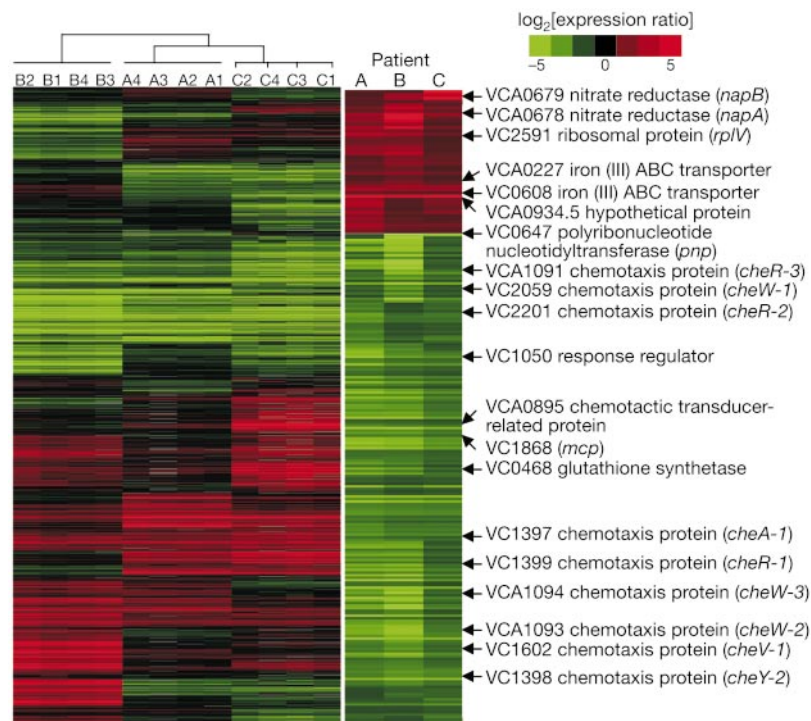


Figure 2 Transcription profile of human-shed *V. cholerae*. Cluster diagram showing the whole-genome expression profile of *V. cholerae* recovered from the stools of three ICDDR patients (left panel). Patient samples are labelled A–C and replicate arrays are indicated by the numbers 1–4. The dendrogram shows tight clustering of technical replicates. The right panel is a cluster diagram of genes showing consistent differential regulation from patients A, B and C. Fold changes relative to *V. cholerae* grown *in vitro* were calculated

using the average values from the quadruplicate arrays shown on the left, and are schematically represented here. Red indicates a minimum twofold increase in expression; green represents a minimum twofold reduction in expression. Representative genes are listed and their relative location indicated by an arrow. See Supplementary Information for a complete list with relative fold changes.

did not meet cut-offs imposed during the course of annotation. 'VCA0934.5' (residues 885680–885886) lies in the region between VCA0933 and VCA0935, is transcribed in the same direction as the upstream VCA0933, and partially overlaps the divergently transcribed VCA0934. VCA0934.5 has a predicted strong Shine–Dalgarno sequence and would code for a protein comprising 69 amino acids that shows no significant similarity to any proteins contained in public databases and bears no known protein motifs. Elucidation of the function of this factor will be of great interest and its discovery emphasizes the importance of analysis of DNA sequences that lie outside of the cut-offs commonly used for genomic annotation in bacteria.

A marked number of genes involved in chemotaxis showed altered expression levels: expression of these genes was repressed in the stool-derived *V. cholerae*. The role of chemotaxis in virulence of *V. cholerae* is unclear^{16,17}, although several genes required for chemotaxis are known to be crucial for expression of cholera toxin during infection¹⁸. Notably, these latter genes, including *cheA-2*, *cheY-3* and *cheZ*, did not show altered expression in the stool samples. Instead, most of the remaining chemotaxis genes were repressed, including genes for all three CheW linker proteins, three out of four CheV linker proteins, all three CheR methyltransferases, and 17 methyl-accepting chemotaxis proteins. Although the stool-derived *V. cholerae* were highly motile as assessed by dark-field and phase-contrast microscopy, many of the gene families that are essential for chemotactic signalling and adaptation to chemoattractant and chemorepellent gradients were repressed. These results suggest that these motile bacteria are non-chemotactic on exiting the human host, and imply that downregulation of the chemotactic response is linked to dissemination from the host and re-entry to the aquatic environment. One possibility is that repression of chemotaxis increases shedding from the gastrointestinal tract. Finally, loss of the chemotactic response in otherwise motile *V. cholerae* was recently shown to markedly increase infectivity in an infant mouse model of cholera¹⁸. It is possible that the motile but non-chemotactic state combined with the unique physiological state of human-shed *V. cholerae* mediates the observed increase in infectivity. Because this transcriptional profiling provides only a 'snap-shot' of gene expression at the moment of dissemination, an important next step will be the elucidation of the proteome of human-shed *V. cholerae*.

Our findings identify the process by which cholera epidemics may be propagated by the human host, and provide insight into gene expression patterns of the disseminating microorganisms. Passage through the human gastrointestinal tract induces a hyper-infectious state, which is perpetuated even after purging into natural aquatic reservoirs. The fact that the human host not only provides a suitable niche for growth but also prepares *V. cholerae* for infection of additional humans has interesting implications for the study of human-to-human spread of other virulent microorganisms. Perhaps other epidemic pathogens are spread in a manner whereby the agent is made more infectious before or during exit from its primary host. Elucidation of the gene expression profiles of disseminating microorganisms should provide new targets for antimicrobial therapy to disable transmission and for vaccine development to prevent infection. □

Methods

Strains, sample collection and competition assays

Strain DSM-V984 was differentially marked at the *lacZ* locus by mobilization of the suicide plasmid pGP704::lacZ and selection on media supplemented with ampicillin and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) as described previously¹⁹. Patients admitted to the ICDDR, Bangladesh, showing clinical manifestations of cholera had rice-water stools pre-screened by dark-field microscopy to verify the presence of *V. cholerae*, and by inhibition of dark-field motility using monoclonal antibodies to determine serogroup and serotype. Freshly shed stool samples were collected in sterile beakers, filtered through cheesecloth, and a portion was frozen at –80 °C for subsequent RNA preparation. The remainder of the stool sample was used in competition assays. Human

studies protocols were reviewed and approved by the Research Review Committee and Ethical Review Committee at the ICDDR, Bangladesh, and by the Institutional Review Board at the Massachusetts General Hospital.

Competition assays were conducted by mixing DSM-V984 grown on LB medium overnight with stool bacteria in a ratio of 1:10 (v/v). A total of 50 µl of a 1,000-fold dilution of this mixture, representing 10⁴–10⁵ colony-forming units (c.f.u.), was administered to 3–5-day-old Swiss Webster mice by gavage as described¹⁴. After 20–24 h, animals were euthanized, the small intestines removed, homogenized and plated on LB medium supplemented with X-gal to allow for subsequent enumeration of LacZ⁺ and LacZ[–] colonies. Outputs ranged from 10³ to 10⁷ c.f.u. per small intestine. Ratios of output bacteria were corrected for deviations in input ratios from 1:1 and represent the competitive index of the strains in question. Competitions with water-incubated stool-derived bacteria were conducted as above with the exception that stool-derived *V. cholerae* were diluted into pond water and were incubated at room temperature for 5 h before mixing with DSM-V984 and inoculation into mice. The pH of the two pond water samples was 7.0–7.5.

Microarray analysis

The construction of the *V. cholerae* DNA microarray will be described in greater detail elsewhere. Briefly, after the completion of the genomic sequence of N16961 by the Institute for Genomic Research, but before annotation of the genome, predicted unique ORFs were found using two separate ORF-finding programs, and a portion of each ORF was amplified by polymerase chain reaction (PCR) and spotted onto poly-L-lysine-coated glass slides¹⁵. The array consists of 5,222 spots representing 3,357 ORFs with an average size of 600 base pairs per PCR product.

Total *V. cholerae* RNA was collected from stool samples and strain DSM-V999 grown overnight *in vitro* using TRIzol reagent (GIBCO-BRL) as described¹⁵, with the exception that RNA was collected on an RNeasy column (QIAGEN) immediately after isopropanol precipitation. DNase treatment to remove DNA contamination was conducted before elution from the column. Quantities of RNA were determined by measuring absorbance at 260/280 nm (*A*₂₆₀/*A*₂₈₀) and integrity was verified by visualization on a 1% agarose gel. Equal concentrations of each test RNA and the common reference RNA (exponentially growing 92A1552) were used for reverse transcription reactions that directly incorporate Cy5- and Cy3-labelled nucleotides into the cDNA as described¹⁵. Labelling reactions were performed in duplicate on two separate days, resulting in quadruplicate arrays for each strain derived from stool samples or grown *in vitro*. The hybridizations and data analyses were performed as described in the main text. Control arrays were also hybridized to identify genes potentially affected by the stool-freezing process, and statistically significant changes that matched patient stool sample regulation were eliminated from subsequent analysis (all microarray data are provided at <http://genome-www5.stanford.edu/microarray/SMD/>).

Received 18 March; accepted 22 April 2002; doi:10.1038/nature00778.

- Cash, R. A. *et al.* Response of man to infection with *Vibrio cholerae*. I. Clinical, serologic, and bacteriologic responses to a known inoculum. *J. Infect. Dis.* **129**, 45–52 (1974).
- Gitelson, S. Gastrectomy, achlorhydria and cholera. *Israel J. Med. Sci.* **7**, 663–667 (1971).
- Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
- Heidelberg, J. F. *et al.* DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477–483 (2000).
- Freter, R., Smith, H. L. & Sweeney, F. J. Enumeration of cholera vibrios in faecal samples. *J. Infect. Dis.* **109**, 31–34 (1961).
- Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
- Freter, R., Smith, H. L. & Sweeney, F. J. An evaluation of intestinal fluids in the pathogenesis of cholera. *J. Infect. Dis.* **109**, 35–42 (1961).
- Miller, V. L. & Mekalanos, J. J. Synthesis of cholera toxin is positively regulated at the transcriptional level by *toxR*. *Proc. Natl Acad. Sci. USA* **81**, 3471–3475 (1984).
- Carroll, P. A., Tashima, K. T., Rogers, M. B., DiRita, V. J. & Calderwood, S. B. Phase variation in *tcpH* modulates expression of the *ToxR* regulon in *Vibrio cholerae*. *Mol. Microbiol.* **25**, 1099–1111 (1997).
- Hase, C. C. & Mekalanos, J. J. TcpP protein is a positive regulator of virulence gene expression in *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **95**, 730–734 (1998).
- DiRita, V. J., Parsot, C., Jander, G. & Mekalanos, J. J. Regulatory cascade controls virulence in *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **88**, 5403–5407 (1991).
- Herrington, D. A. *et al.* Toxin, toxin-coregulated pili, and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. *J. Exp. Med.* **168**, 1487–1492 (1988).
- Merrell, D. S., Hava, D. L. & Camilli, A. Identification of novel factors involved in colonization and acid tolerance of *Vibrio cholerae*. *Mol. Microbiol.* **43**, 1471–1491 (2002).
- Merrell, D. S. & Camilli, A. The *cadA* gene of *Vibrio cholerae* is induced during infection and plays a role in acid tolerance. *Mol. Microbiol.* **34**, 836–849 (1999).
- Schoolnik, G. K. *et al.* Whole genome DNA microarray expression analysis of biofilm development by *Vibrio cholerae* O1 El Tor. *Methods Enzymol.* **336**, 3–18 (2001).
- Richardson, K. Roles of motility and flagellar structure in pathogenicity of *Vibrio cholerae*: analysis of motility mutants in three animal models. *Infect. Immun.* **59**, 2727–2736 (1991).
- Gardel, C. L. & Mekalanos, J. J. Alterations in *Vibrio cholerae* motility phenotypes correlate with changes in virulence factor expression. *Infect. Immun.* **64**, 2246–2255 (1996).
- Lee, S. H., Butler, S. M. & Camilli, A. Selection for *in vivo* regulators of bacterial virulence. *Proc. Natl Acad. Sci. USA* **98**, 6889–6894 (2001).
- Camilli, A. & Mekalanos, J. J. Use of recombinase gene fusions to identify *Vibrio cholerae* genes induced during infection. *Mol. Microbiol.* **18**, 671–683 (1995).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank E. A. Joyce, M. K. Waldor and A. L. Sonenshein for comments and discussion of the manuscript, C. C. Kim for help with statistical analysis, and S. Falkow and J. J. Mekalanos for critical discussions. This research was supported by grants from the National Institutes of Health. The ICDDR at Bangladesh is supported by agencies and countries that share its concern for the health problems of the developing world.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.C. (e-mail: Andrew.Camilli@Tufts.edu).

Origin of GABAergic neurons in the human neocortex

Kresimir Letinic, Roberto Zoncu & Pasko Rakic

Yale University School of Medicine, Section of Neurobiology, New Haven, Connecticut 06510, USA

The mammalian neocortex contains two major classes of neurons, projection and local circuit neurons^{1–4}. Projection neurons contain the excitatory neurotransmitter glutamate, while local circuit neurons are inhibitory, containing GABA^{2,4}. The complex function of neocortical circuitry depends on the number and diversity of GABAergic (γ -aminobutyric-acid-releasing) local circuit neurons^{1–3}. Using retroviral labelling in organotypic slice cultures of the embryonic human forebrain, we demonstrate the existence of two distinct lineages of neocortical GABAergic neurons. One lineage expresses *Dlx1/2* and *Mash1* transcription factors, represents 65% of neocortical GABAergic neurons in humans, and originates from *Mash1*-expressing progenitors of the neocortical ventricular and subventricular zone of the dorsal forebrain. The second lineage, characterized by the expression of *Dlx1/2* but not *Mash1*, forms around 35% of the GABAergic neurons and originates from the ganglionic eminence of the ventral forebrain. We suggest that modifications in the expression pattern of transcription factors in the forebrain may underlie species-specific programmes for the generation of neocortical local circuit neurons^{5–11} and that distinct lineages of cortical interneurons may be differentially affected in genetic and acquired diseases of the human brain.

Classical studies, as well as modern retroviral lineage analyses, concur that projection neurons, in all species examined, originate in the ventricular zone (VZ) of the dorsal telencephalon and migrate radially across the intermediate zone (IZ) to the overlying cortical plate^{6,7,10–14}. However, recent retroviral lineage and chimera studies in rodents have demonstrated that local circuit neurons originate in the subcortical ganglionic eminence (GE) of the ventral telencephalon and migrate tangentially through the IZ into the cerebral cortex^{5–11,15–17}. Furthermore, mice lacking the transcription factors *Dlx* or *Mash1* that are expressed in the GE but not in the VZ of the dorsal telencephalon lack most cortical GABAergic neurons, confirming their origin from the ventral telencephalon^{10,18,19}. Because GABAergic neurons increase in number, complexity and proportion relative to projection neurons during primate evolution^{1,2,4,20,21} and are implicated in a host of neurological disorders, we examined their origin by lineage analysis in slices of embryonic human forebrain tissue.

The developmental period analysed ranged from 10 to 25 weeks of gestation, which corresponds to the major wave of generation and migration of cortical neurons in the human brain²². To

determine the pattern of migration of GABAergic neurons, we first immunostained cortical sections for GABA, *Dlx1/2* homeo-domain proteins and neuron-specific β -tubulin/TuJ1 antigen and analysed immunoreactive cells with migratory morphologies. We found that immunoreactivity for all three antigens (GABA/*Dlx*/TuJ1) defined a population of mostly non-radially oriented migratory-like neurons within the ventricular and subventricular zone (VZ/SVZ) (Fig. 1a, b). Numerous migratory-like, radially oriented GABA/*Dlx*-immunoreactive cells were present in the intermediate zone (IZ), situated between the proliferative VZ/SVZ and the cortical plate (Fig. 1c).

We next performed a direct analysis of cell migration in organotypic slices of the embryonic cerebral cortical wall. Crystals of DiI were placed into the VZ/SVZ and the distribution of migrating cells examined after 36–48 hours in culture (Fig. 1d, e). At 10 weeks of gestation 95% of labelled cells analysed ($n = 180$ from 21 slices) migrated radially in the VZ/SVZ and the IZ. From 13 weeks of gestation the incidence of non-radially migrating cells and their proportion in the total population of migrating cells significantly increased, although we did not observe an overall increase in the magnitude of cell migration (that is, in the number of cells observed away from a DiI crystal per slice). Thus, in the VZ/SVZ of specimens that were 14 and 24 weeks old, 60–65% of labelled cells analysed ($n = 255$ cells from 30 slices at 14 weeks of gestation; $n = 320$ cells from 28 slices at 24 weeks of gestation) migrated non-radially, even though in the IZ, around 75% of labelled cells analysed ($n = 150$ per fetal stage) migrated radially (Fig. 1d, e).

To demonstrate directly non-radial migration of neuronal cells within the VZ/SVZ, we designed a DNA construct encoding a yellow fluorescent protein (EYFP) under the control of neuron-specific Ta1 promoter (Fig. 1f). This construct labels neurons exclusively (Fig. 1g, h) and enables time-lapse analysis of their migration (Fig. 1i). We have observed an extensive non-radial migration of EYFP-positive cells in the VZ/SVZ of neocortical slice cultures (for example, see Fig. 1i). Electron microscopic analysis confirmed the existence of a network of non-radially migrating TuJ1/GABA-positive neurons within the VZ/SVZ and suggested that these cells rely on neurophilic-homotypic guidance (Fig. 1j). Moreover, when neocortical VZ/SVZ explants were cultured in Matrigel, a massive neurophilic migration of TuJ1/GABA/*Dlx*-positive neurons was noted out of the explants (Fig. 1k, l), whereas only gliophilic migration along radial glial fibres occurred out of IZ explants (Fig. 1m). Thus, both morphological and *in vitro* data together indicate that GABAergic neurons migrate in non-radial, neurophilic fashion within the VZ/SVZ, but switch to radial, gliophilic mode in the IZ, after exiting the VZ/SVZ.

We observed in the developing human cerebral wall a distinct subpopulation of GABAergic neurons, that in addition to *Dlx1/2* also expresses *Mash1* (Fig. 2a–d). Thus, at 25 weeks of gestation, $65.1 \pm 1.6\%$ of GABA/*Dlx1/2*-positive cortical plate neurons analysed ($n = 600$) expressed *Mash1*. Furthermore, about two-thirds of migrating GABAergic neurons in the VZ/SVZ analysed ($n = 600$), belonged to this class ($70.2 \pm 3.3\%$ at 14 weeks and $61.8 \pm 3.0\%$ at 24 weeks of gestation). Strikingly, our analysis also revealed that 8% of M-phase VZ progenitors and 85% of M-phase SVZ progenitors analysed ($n = 180$ for VZ and 240 for SVZ) expressed *Mash1* (Fig. 2e–g), thus strongly suggesting that *Mash1*-positive GABAergic neurons originate in the dorsal forebrain proliferative zone.

To test this possibility, we exposed slices of the embryonic human cerebral cortex to a replication-incompetent retrovirus encoding green fluorescent protein (GFP) under a constitutive promoter. The slices did not contain the GE, thus excluding the possibility of labelling cortical migrating cells that may originate in the GE. In the VZ/SVZ of the dorsal telencephalon, GFP-positive clones were observed 18–24 h after the infection and their position periodically recorded every 2–6 h for two to three days (or every 20–30 min for a few hours in some cases). The infections resulted in one or two

labelled cells per slice that would later form small clones. We included in our analysis only those slices where the first appearing labelled cells were widely separated, which was most often the case in large (≤ 1 cm) slices of the human cortex. This enabled an unambiguous identification of clonal boundaries. We have focused on the clones containing non-radially migrating cells in the VZ/SVZ. The retrovirus integrates randomly in one of the daughter cells after infection, so it may label either the progenitor cell or the postmitotic neuron. As expected, one division (24 h) after infection, labelled cells ($n = 114$) were single and time-lapse analysis revealed that they were either dividing progenitors with a large (10–12 μm in diameter) and round cell body and one or two processes (Fig. 3a), or postmitotic migrating cells with a smaller oval soma (4–6 μm in diameter) and either a single leading process (Fig. 3b), or two processes—a leading and a trailing process. Forty-five per cent of retrovirally labelled cells represented postmitotic migrating cells that did not divide during the period of observation, after their initial appearance (Fig. 3e). In contrast, 52% of retrovirally labelled cells divided to form two-cell clones and 3% divided to form three-cell clones. Within the two-cell clones, 82% included one progenitor-like cell and one migratory-like cell, suggesting that the initial

progenitor divided asymmetrically to produce another progenitor and a migrating cell (Fig. 3a and c). The remaining two-cell clones contained two migratory-like cells, suggesting that the initial progenitor divided symmetrically to give rise to two migrating cells (Fig. 3d). Migrating cells in the above analysed cortical clones expressed GABA (19/21) and *Dlx* (18/18) (Fig. 3f, h, i). Furthermore, *Mash1* was expressed both by the dividing progenitors and migrating cells (22/23) (Fig. 3k–m). In a series of parallel experiments, we retrovirally labelled cells migrating out from the dorso-pallial VZ/SVZ explants cultured in Matrigel and showed that they express *Dlx* and differentiate into GABAergic neurons (Fig. 3g and j).

In rodents, most, or all neocortical GABAergic cells originate in the GE^{5–11,15–17}. We examined whether in the human brain, *Mash1*-negative GABAergic neurons also originate in the GE. First, we established that migrating GE neurons in immunostained sections of human forebrain expressed only *Dlx1/2* but not *Mash1* proteins (Fig. 4a–c). Although *Mash1* was expressed by the progenitors in the GE, it was not detectable in any of the postmitotic neurons leaving this proliferative compartment. The molecular phenotype of the GE-derived migrating neurons was confirmed in GE explants

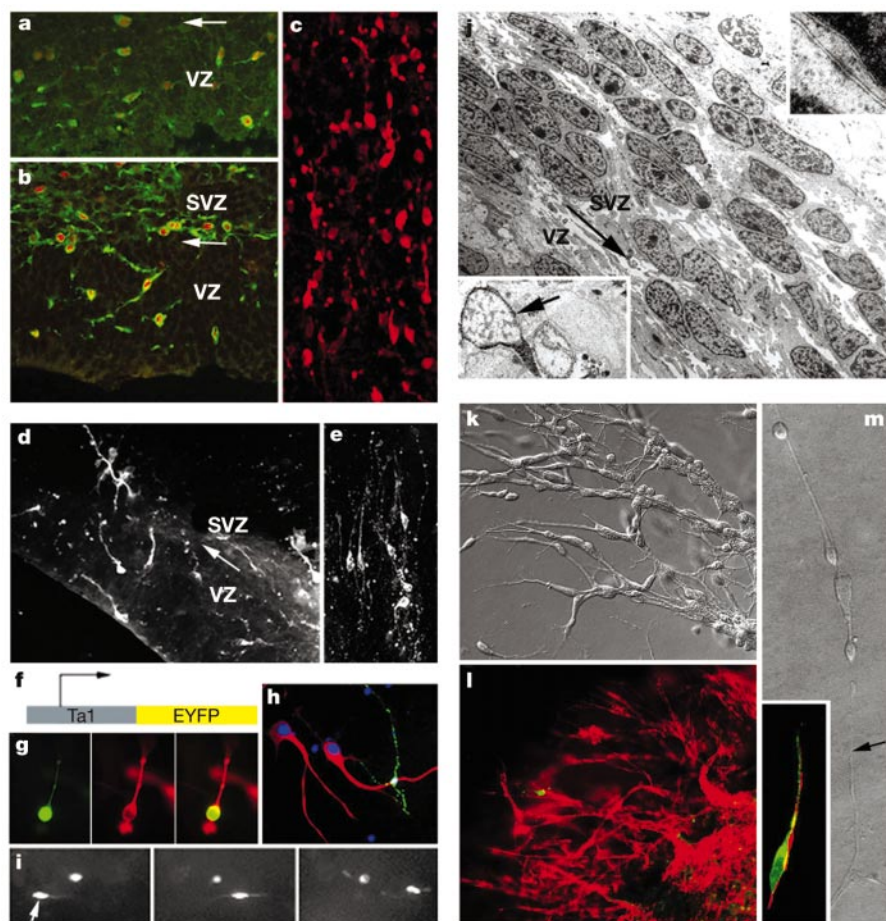


Figure 1 Mode of migration of GABAergic neurons in the fetal human neocortex. **a, b**, Migrating GABAergic neurons in the VZ and SVZ at 13 weeks of gestation stained for GABA (green) and *Dlx2* (red, nuclear) (**a**), or for *TuJ1* (green) and *Dlx2* (red) (**b**). The arrow marks the VZ/SVZ border in **a, b, d** and **j**. **c**, Radially migrating GABA-positive cells in the IZ at 18 weeks. **d**, Dil-labelled, non-radially migrating cells in the cortical VZ/SVZ at 14 weeks. **e**, Dil-labelled radially migrating cells in the IZ at 14 weeks. **f**, A schematic drawing of Ta1-EYFP plasmid. **g, h**, EYFP fluorescence (green) is present only in *TuJ1*-positive neurons (red) (**g**), but not GFAP-positive glial cells (red); blue is a nuclear staining (**h**). **i**, A time lapse of a tangentially migrating EYFP-positive neuron (arrow) in the cortical SVZ. Time difference among the frames is 45 min. **j**, Electron micrograph showing chains of

non-radially migrating neurons at the VZ/SVZ border. The upper inset shows a zonula-adherens-like contact between membranes of apposed neurons. The lower inset shows GABA-immunoreactivity in a non-radially migrating SVZ neuron (arrow). **k, l**, High-power differential interference contrast (DIC) image (**k**) and GABA (red)/GFAP (green) double staining (**l**) of migrating neuronal chains coming out from cortical VZ/SVZ explants. Notice an absence of radial glial fibres among migrating chains. **m**, High-power DIC image of a radial glial fibre (arrow) and migrating neurons attached to it, coming out of cortical IZ explants; no neurophilic chain migration is present in the case of IZ explants, but only radial migration along radial glia. The inset shows a GABA-positive neuron (green) attached to one such GFAP-positive radial glial fibre (red).

cultured in VitroGel, or Matrigel. At all developmental stages, migrating cells were exclusively TuJ1-positive, BrdU-negative postmitotic neurons. Progenitors that are revealed by their BrdU incorporation were not seen migrating out of the explants. Importantly, migrating postmitotic neurons were *Dlx1/2*-positive, but never *Mash1*-positive (Fig. 4d, e). Finally, by placing DiI crystals in the GE of the forebrain slices we observed migration of labelled cells into the developing neocortex. At 10 weeks of gestation, only a few cells (0–2 per slice) migrated into the IZ, but not VZ of the dorsal telencephalon, whereas at later stages a robust stream of migration spread through the VZ/SVZ (Fig. 4f). These data suggest that *Mash1*-negative neocortical GABAergic neurons originate in the GE, as they do in mice^{5–11,15–17}.

The present study identifies a distinct population of *Mash1*-expressing progenitors for GABAergic neurons (Fig. 4g) in the human dorsopallial proliferative zones. We suggest that *Mash1* transcription factor may induce GABAergic neuronal production

in the human dorsopallial VZ/SVZ by upregulating *Dlx* genes. In support of this hypothesis, an ectopic *Mash1* expression in rodents is sufficient to upregulate *Dlx1/2* in neocortical neurons²³, whereas an ectopic *Dlx* expression may induce the GABAergic phenotype in the same cells¹⁰. We found *Mash1* expression only in those cortical plate neurons that express GABA and *Dlx*, suggesting that *Mash1*-positive progenitors do not give rise to pyramidal neurons. Even though rodent SVZ may give rise to some cortical neurons²⁴, most retroviral labelling studies indicate that SVZ is predominantly a gliogenic compartment²⁵. This is in contrast to a massive neuro-

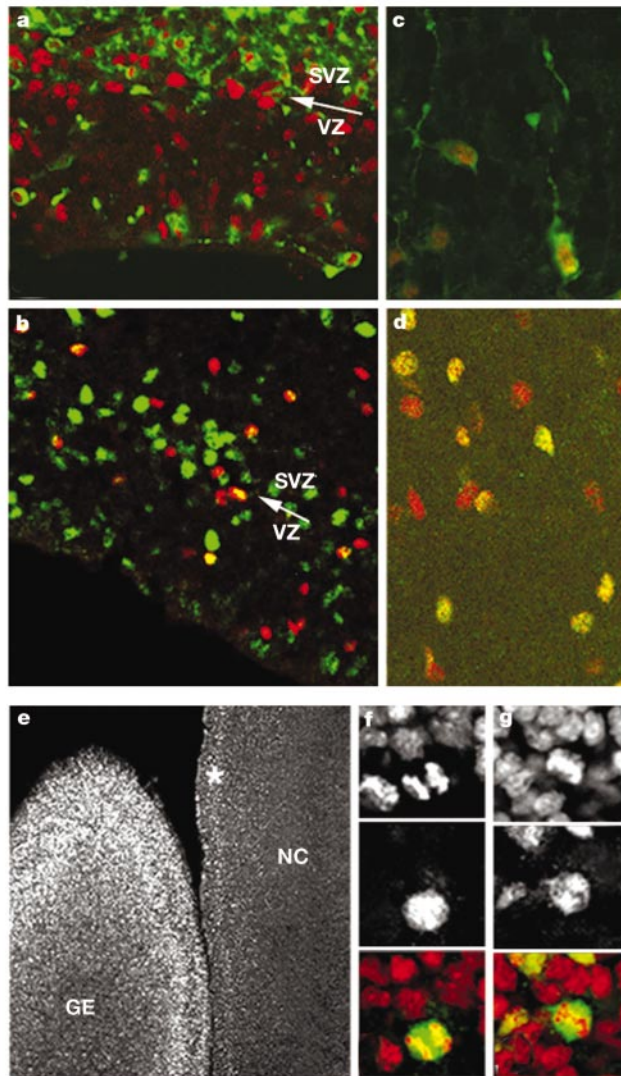


Figure 2 *Mash1*-expression in neocortical GABAergic neurons and VZ/SVZ progenitors. **a, b**, TuJ1 (green)/*Mash1* (red) (**a**), or *Mash1* (green)/*Dlx2* (red) (**b**) double staining of the cortical VZ/SVZ at 13 weeks. **c**, GABA (green) and *Dlx2* (red) staining of the cortical plate neurons at 18 weeks of gestation. **d**, *Mash1* (green) and *Dlx2* (red) staining of the cortical plate at 25 weeks. **e**, *Mash1* staining of the forebrain at 14 weeks (asterisk marks the neocortical VZ/SVZ). GE, ganglionic eminence; NC, neocortex. **f, g**, *Mash1* (green) in M-phase cells of the VZ (**f**) and SVZ (**g**) labelled by propidium iodide (red).

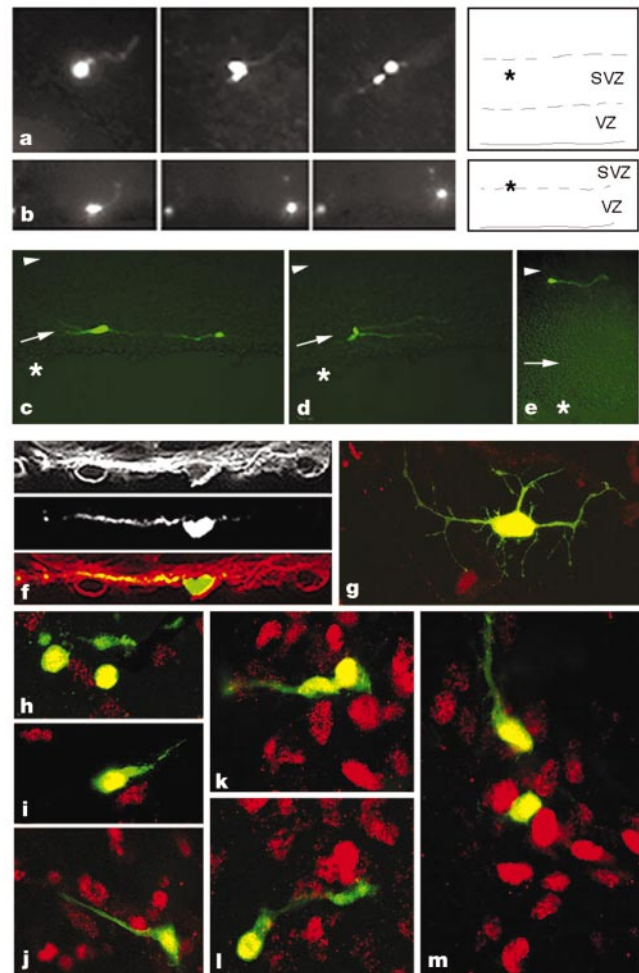


Figure 3 Neocortical lineage of GABAergic neurons and their progenitors revealed by a retroviral analysis in living slices. **a, b**, Time-lapse series of a dividing neocortical SVZ progenitor (**a**), or a newborn, non-radially migrating cell (**b**); schematics on the right illustrate the locations of these clones. Difference among frames is 30 min. **c**, A two-cell clone at the VZ/SVZ border, containing a progenitor-like cell (left) and a migrating cell (right). **d**, Two-cell clone at the VZ/SVZ border, containing two migrating cells. **e**, Single cell clone—a migrating cell in the SVZ. In **c–e**, the asterisk marks the lower VZ edge, the arrow marks the VZ/SVZ border, and the arrowhead marks the upper SVZ border. **f**, Retrovirally labelled migrating neuron in the neocortical SVZ expressing GABA. The top image is GABA staining, the middle one shows an EGFP-positive cell and the bottom one is a double exposure. **g**, Retrovirally labelled EGFP-positive neuron that migrated out of a cortical VZ/SVZ explant and differentiated, expressing GABA (red) (visible as yellow in double exposure). **h**, SVZ clone composed of two migrating neurons expressing *Dlx2* (red in **h–j**) in their nuclei. **i**, Single-cell SVZ clone—*Dlx2*-expressing migrating neuron. **j**, Retrovirally labelled neuron that migrated out of a cortical VZ/SVZ explant, positive for *Dlx2*. **k**, SVZ clone containing a dividing progenitor positive for *Mash1* (red in **k–m**). One of the daughters resembles a non-radially migrating neuron. **l**, Single-cell SVZ clone—*Mash1*-positive migrating neuron. **m**, Two-cell clone positive for *Mash1*; the upper cell resembles a radially migrating neuron.

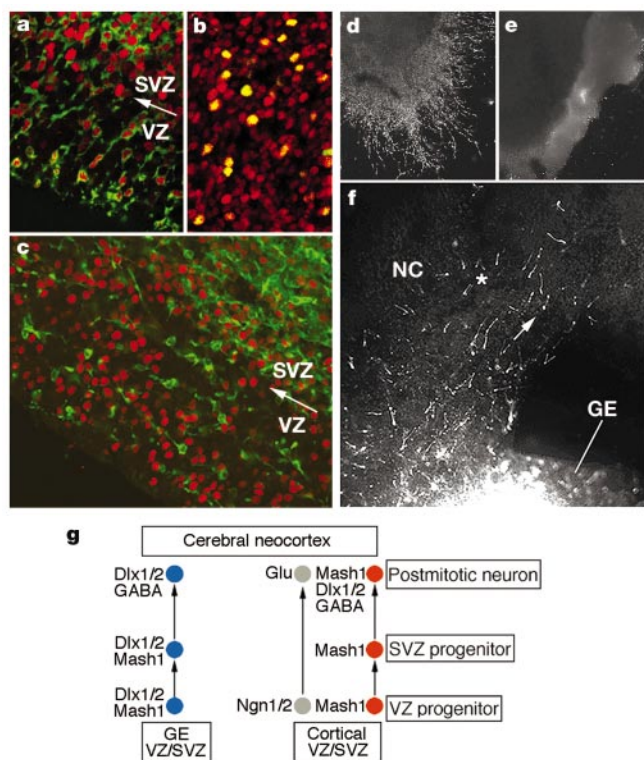


Figure 4 Neocortical GABAergic neurons originating in the GE. **a**, Migrating neurons TuJ1-positive (green) in the GE express DLX2 (red). **b**, Double staining for Mash1 (green) and DLX2 (red) in the SVZ of the GE, showing that only a small subset of DLX-positive cells also expresses Mash1 (visible as yellow). **c**, TuJ1-positive (green) migrating neurons of the GE do not express Mash1 (red). **d**, **e**, Cells migrating out from the GE explant are DLX2-positive (**d**), but Mash1-negative (**e**). **f**, A low-power of a slice from a 15-week-old brain, showing many Dil-labelled cells migrating from the GE into the cortex. GE, ganglionic eminence; NC, neocortex. **g**, A model illustrating the origin of neocortical neurons in the human brain. The separate lineages of GABAergic neurons that derive from either the GE (blue), or the neocortical VZ/SVZ progenitors (orange) are indicated, as well as the molecular phenotype of progenitors and postmitotic neurons. The origin of projection, glutamatergic (Glu) neurons from the neocortical VZ progenitors (presumably expressing neurogenin that is expressed by these cell in rodents) is also indicated (grey).

genesis by SVZ progenitors observed in the present study. However, we cannot exclude the possibility that some of Mash1-expressing progenitors in the neocortical VZ/SVZ in fact arrived from the GE at earlier embryonic stages than those that were analysed here and then continued to divide locally. Nevertheless, our data strongly suggest that Mash1-positive neocortical GABAergic neurons originate (are born) in the VZ/SVZ of the dorsal telencephalon as a distinct neuronal stem cell lineage. Another example of a neuronal lineage apparently introduced during primate brain evolution is a population of GE-derived GABAergic neurons migrating to the selected thalamic nuclei³². The identification of the neocortical origin of GABAergic neurons in the developing human cerebrum may provide insight into mechanisms operating during primate brain evolution^{1,2,26,27}. Furthermore, revealing distinct lineages of GABAergic neurons may help to understand the pathogenesis of congenital and acquired neurological disorders related to defects of separate classes of local circuit neurons^{2,28–30}.

Methods

Immunohistochemistry and electron microscopy

Post-mortem brain tissue blocks were fixed in 4% paraformaldehyde and 0.1% glutaraldehyde and either cryoprotected and cut on a cryostat for light microscopy or cut on a vibratome for electron microscopy. Primary antibodies and dilutions were as follows: TUJ1 monoclonal (Becton-Dickinson, 1:500); anti-GABA polyclonal (SIGMA, 1:10,000);

anti-Mash1 monoclonal and polyclonal (gift of J. E. Johnston; 1:1 and 1:1,000, respectively); anti-DLX1 and anti-DLX2 polyclonals (gift of J. L. R. Rubenstein; 1:20 and 1:200, respectively). The sections were blocked, incubated with primary antibodies overnight at 4 °C and then with appropriate secondary antibodies overnight at 4 °C (Jackson). Some tissue blocks were fixed in 2% paraformaldehyde and 2.5% glutaraldehyde, osmicated in 1% osmium tetroxide and embedded in Durcupan (Fluka Chemie). Selected blocks were cut serially into ultrathin sections on a Leica/Reichert-Yung Ultramicrotome, stained with lead citrate and uranyl acetate and examined with a JEOL transmission electron microscope. Vibratome sections were processed for immunoperoxidase histochemistry, osmicated, embedded flat in Durcupan and processed as above.

Slice and explant cultures

Post-mortem human brain tissue (from specimens ranging from 10 to 25 weeks of gestation) was obtained from the Albert Einstein College of Medicine, New York State licensed Human Fetal Repository, approved by the institutional board. Research protocols were approved by the Human Investigation Committee at Yale University School of Medicine. We have used embryonic organotypic slice preparations which were shown to reproduce faithfully both the timing and mode of cell proliferation and migration in the cerebral wall for up to at least 72 h (ref. 31). Shortly afterwards, brains were put in HEPES-buffered MEM (GIBCO) with L-glutamine (GIBCO) on ice. Dissected blocks of cerebral hemispheres were cut using a McIlwan tissue chopper (Mickle Lab Engineering Co.) into 300- μ m-thick slices. Slices were cultured on Millicell inserts (Millipore), placed onto 35-mm Petri dishes and incubated in Neurobasal medium with L-glutamine, B27, N2 and antibiotics (GIBCO) at 37 °C with 5% CO₂. DiI (1,1'-diiododecyl-3,3,3'-tetramethylindocarbocyanine perchlorate) crystals (Molecular Probes) were placed using insert pins under a dissection microscope. After culturing for up to 72 h, slices were fixed in 4% paraformaldehyde, immunostained and analysed by laser confocal microscopy. In some cases, small explants of the GE, or neocortical VZ/SVZ and IZ were isolated, embedded in Vitrogen, or Matrigel in 35-mm dishes with glass bottoms (MatTek), cultured and finally processed as slices.

Plasmid construction and transfection

Tal promoter was subcloned into EYFP promoterless plasmid (Clontech). Transfection of dissociated cortical cell cultures as well as slices was performed using the Fugene transfection kit (Roche). Expression of EYFP was evident 24 h after transfection.

Retroviral production and infection

A replication-incompetent, amphotropic retroviral vector was produced by transfecting the pLEGFP-N1 vector (CLONTECH) into Phoenix amphotropic packaging cell line (Nolan laboratory, Stanford University), by calcium phosphate precipitation. Viral supernatant was collected 36 h after transfection, concentrated 50 times by centrifugation, aliquoted and stored at -80 °C until use. The virus titre, determined on NIH3T3 fibroblast, was 5×10^5 c.f.u. ml⁻¹. Helper virus was not detected. Slices and explants were infected by adding 20 μ l of retroviral preparation to 800 μ l of Neurobasal/B27/N2 serum containing 8 mg ml⁻¹ polybrene and incubating for 1 h.

Received 3 November 2001; accepted 21 February 2002; doi:10.1038/nature00779.

- Ramon y Cajal, S. R. *Histologie du Système Nerveux de l'Homme et des Vertébrés* (Maloine, Paris, 1911).
- Rakic, P. Local circuit neurons. *Neurosci. Res. Prog. Bull.* **13**, 1–399 (1975).
- Gupta, A., Wang, Y. & Markram, H. Organizing principle for diversity of GABAergic interneurons and synapses in the neocortex. *Science* **287**, 273–278 (2000).
- Fairen, A., DeFelipe, J. & Regidor, J. in *Cellular Components of the Cerebral Cortex* (eds Peters, A. & Jones, E. G.) 201–253 (Plenum, New York, 1984).
- Marin, O., Yaron, A., Bagri, A., Tessier-Lavigne, M. & Rubenstein, J. L. R. Sorting of striatal and cortical interneurons regulated by semaphorin-neuropilin interactions. *Science* **293**, 872–875 (2001).
- Tan, S. S. *et al.* Separate progenitors for radial and tangential cell dispersion during development of the cerebral neocortex. *Neuron* **21**, 295–304 (1998).
- Ware, M. L., Tavazoie, S. F., Reid, C. B. & Walsh, C. A. Coexistence of widespread clones and large radial clones in early embryonic ferret cortex. *Cereb. Cortex* **9**, 636–645 (1999).
- de Carlos, J. A., López-Mascaraque, L. & Valverde, F. Dynamics of cell migration from the lateral ganglionic eminence in the rat. *J. Neurosci.* **16**, 6146–6156 (1996).
- Powel, E. M., Mara, W. M. & Levitt, P. Hepatocyte growth factor/scatter factor is mitogen for interneuron migrating from the ventral to dorsal telencephalon. *Neuron* **30**, 1–20 (2001).
- Anderson, S., Mione, M., Yun, K. & Rubenstein, J. L. R. Differential origins of neocortical projection and local circuit neurons: Role of *Dlx* genes in neocortical interneuronogenesis. *Cereb. Cortex* **9**, 646–654 (1999).
- Lavdas, A. A., Grigoriou, M., Pachnis, V. & Parnavelas, J. G. The medial ganglionic eminence gives rise to a population of early neurons in the developing cerebral cortex. *J. Neurosci.* **19**, 7881–7888 (1999).
- Rakic, P. Mode of cell migration to the superficial layers of fetal monkey neocortex. *J. Comp. Neurol.* **145**, 61–84 (1972).
- Luskin, M. B., Pearlman, A. L. & Sanes, J. R. Cell lineage in the cerebral cortex of the mouse studied *in vivo* and *in vitro* with recombinant retrovirus. *Neuron* **1**, 635–647 (1988).
- Kornack, D. R. & Rakic, P. Radial and horizontal deployment of clonally related cells in the primate neocortex: Relationship to distinct mitotic lineages. *Neuron* **15**, 311–321 (1995).
- Neyt, C., Welch, M., Langston, A., Kohtz, J. & Fishell, G. A short-range signal restricts cell movement between telencephalic proliferative zones. *J. Neurosci.* **17**, 9194–9203 (1997).
- O'Rourke, N. A., Chenn, A. & McConnell, S. K. Postmitotic neurons migrate tangentially in the cortical ventricular zone. *Development* **124**, 997–1005 (1997).
- He, W., Ingraham, C., Rising, L., Goderie, S. & Temple, S. Multipotent stem cells from the mouse basal forebrain contribute GABAergic neurons and oligodendrocytes to the cerebral cortex during embryogenesis. *J. Neurosci.* **21**, 8854–8862 (2001).
- Casasosa, S., Fode, C. & Guillemot, F. Mash1 regulates neurogenesis in the ventral telencephalon. *Development* **126**, 525–534 (1999).

19. Horton, S., Meredith, A., Richardson, J. A. & Johnson, J. E. Correct coordination of neuronal differentiation events in ventral forebrain requires the bHLH factor MASH1. *Mol. Cell. Neurosci.* **14**, 355–369 (1999).
20. Hendry, S. H. & Carder, R. K. Neurochemical compartmentation of monkey and human visual cortex; similarities and variations in calbindin immunoreactivity across species. *Vis. Neurosci.* **10**, 1109–1120 (1993).
21. Jones, E. G. GABAergic neurons and their role in cortical plasticity in primates. *Cereb. Cortex* **3**, 361–372 (1993).
22. Sidman, R. L. & Rakic, P. Neuronal migration, with special reference to developing human brain: A review. *Brain Res.* **62**, 1–35 (1973).
23. Fode, C. et al. A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes Dev.* **14**, 67–80 (2000).
24. Tarabykin, V., Stoykova, A., Usman, N. & Gruss, P. Cortical upper layer neurons derive from the subventricular zone as indicated by *Svet1* gene expression. *Development* **128**, 1983–1993 (2001).
25. Kakita, A. & Goldman, J. Patterns and dynamics of SVZ cell migration in the postnatal forebrain: monitoring living progenitors in slice preparations. *Neuron* **23**, 461–472 (1999).
26. Preuss, T. Taking the measure of diversity: comparative alternatives in the model-animal paradigm in cortical neuroscience. *Brain Behav. Evol.* **55**, 287–299 (2000).
27. Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).
28. Gleeson, J. G. & Walsh, C. A. Neuronal migration disorders: From genetic diseases to developmental mechanisms. *Trends Neurosci.* **23**, 352–359 (2000).
29. Jones, E. G. Cortical development and thalamic pathology in schizophrenia. *Schizophr. Bull.* **23**, 483–501 (1997).
30. Lewis, D. A. GABAergic local circuit neurons and prefrontal cortical dysfunction in schizophrenia. *Brain Res. Rev.* **31**, 270–276 (2000).
31. Haydar, T. F., Bambrick, L. L., Kruger, B. K. & Rakic, P. Embryonic organotypic slice cultures for analysis of proliferation, cell death and migration in the cerebral wall. *Brain Res. Protocols* **4**, 425–437 (1999).
32. Letinic, K. & Rakic, P. Telencephalic origin of human thalamic GABAergic neurons. *Nature Neurosci.* **4**, 931–936 (2001).

Acknowledgements

We thank J. E. Johnson for providing Mash1 antibodies and J. L. R. Rubenstein and S. Anderson for providing *Dlx1/2* antibodies. We also thank F. Miller for providing Ta1-LacZ plasmid. We are grateful to all colleagues in the laboratory of P. R. for their advice and comments on the manuscript.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.R. (e-mail: pasko.rakic@yale.edu).

Regulation of AMPA receptor lateral movements

Aren J. Borgdorff & Daniel Choquet

CNRS UMR 5091, Université de Bordeaux 2, Bordeaux 33077, France

An essential feature in the modulation of the efficacy of synaptic transmission is rapid changes in the number of AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors at post-synaptic sites on neurons^{1–4}. Regulation of receptor endo- and exocytosis has been shown to be involved in this process^{5–14}. Whether regulated lateral diffusion of receptors in the plasma membrane also participates in receptor exchange to and from post-synaptic sites remains unknown. We analysed the lateral mobility of native AMPA receptors containing the glutamate receptor subunit GluR2 in rat cultured hippocampal neurons, using single-particle tracking and video microscopy. Here we show that AMPA receptors alternate within seconds between rapid diffusive and stationary behaviour. During maturation of neurons, stationary periods increase in frequency and length, often in spatial correlation with synaptic sites. Raising intracellular calcium, a central element in synaptic plasticity, triggers rapid receptor immobilization and local accumulation on the neuronal surface. We suggest that calcium influx prevents AMPA receptors from diffusing, and that lateral receptor diffusion to

and from synaptic sites acts in the rapid and controlled regulation of receptor numbers at synapses.

We measured the lateral mobility of native GluR2-containing AMPA receptors in the plasma membrane of rat cultured hippocampal neurons using 0.5- μ m latex beads coated with an antibody against the extracellular domain of GluR2. Each bead was held in contact with the surface of a neurite with laser tweezers for 5–10 s to allow binding to GluR2. After being released from the tweezers, the bead trajectory was followed by video microscopy for 200 s at 25 Hz with a spatial resolution of 5–10 nm, thus reflecting GluR2 movement (Fig. 1a and Supplementary Information).

At all developmental stages studied, AMPA receptors alternated between periods of rapid diffusive movement (blue in Fig. 1a) and periods of slow diffusion confined within a sub-micrometre area (red). Switching between the two types of movement occurred abruptly (within one video frame, or 40 ms). We divided receptor trajectories into episodes expressing one of these behaviours using a confinement index^{15,16} as discriminator (Fig. 1b). Periods with a low confinement index relate to free diffusive movement, whereas a high confinement index reflects slow and spatially restricted movement.

During neuronal maturation, the mean diffusion coefficient of AMPA receptor movement significantly decreased, the main drop occurring at 4–6 days *in vitro* (DIV) (Fig. 1c). The percentage of time spent in the confined state strongly increased as neurons matured. This was due to a 5-fold increase in the mean duration of the confined state and a 1.5-fold decrease in the mean duration of

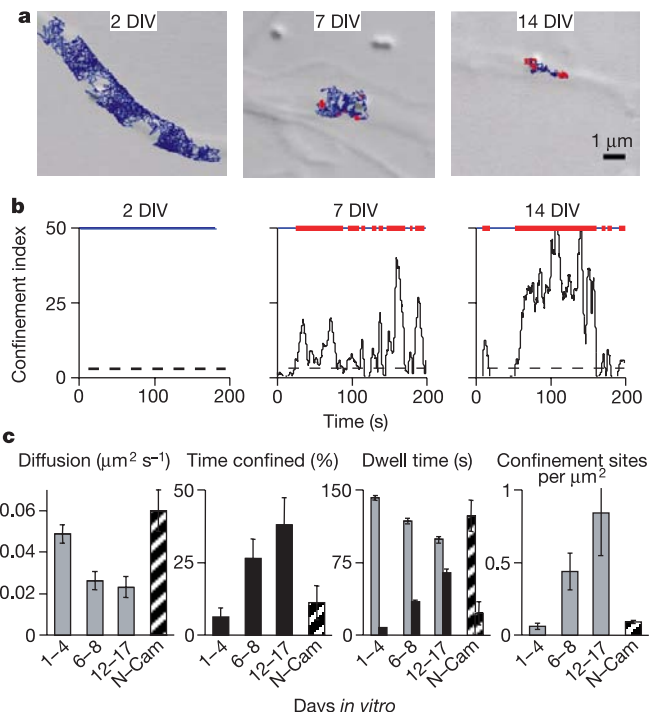


Figure 1 Lateral mobility of GluR2 decreases during neuronal maturation. **a**, Trajectories of GluR2-bound beads recorded for 200 s on the neurites of hippocampal neurons cultured for the indicated number of days (DIV), overlaid on the sample image. Confined and diffusive movement are shown respectively in red and blue. Scale bar, 1 μ m. **b**, Confinement index versus time for the trajectories in **a**. The dashed line shows the threshold for confined and diffusive episodes, as indicated above the plots (colour as in **a**). **c**, Histograms of mobility parameters for GluR2 (plain bars) and N-Cam (hatched bars). The diffusion coefficient, percentage of time spent in confined state, dwell time for diffusive (grey bars) and confined (black bars) episodes, and number of confinement zones divided by the spatial envelope (μ m²) of the trajectory are shown. 1–4 DIV, $n = 31$; 6–8 DIV, $n = 30$; 12–17 DIV, $n = 21$. N-Cam: 8–10 DIV, $n = 15$.

the diffusive state. This relates to the increased number of confinement sites per area of neurite explored by the receptors. By contrast, the mean diffusion coefficient during diffusive and confined periods only slightly decreased during maturation. For diffusive periods, values were (in $\mu\text{m}^2\text{s}^{-1}$, $\times 10^{-2}$): 1–4 DIV, 12 ± 0.9 ; 6–8 DIV, 8 ± 0.8 ; and 12–17 DIV, 6 ± 1 (mean $\pm$ s.e.m.; $n = 31, 30, 21$, respectively). These values were about 30-fold higher than the diffusion coefficient during confined periods, with values of respectively 0.4 ± 0.08 , 0.2 ± 0.05 and 0.2 ± 0.03 (in $\mu\text{m}^2\text{s}^{-1}$, $\times 10^{-2}$). Together, these data show that reversible immobilization periods of AMPA receptors increase in length and frequency during neuronal maturation, in parallel with synaptogenesis¹⁷. For comparison, we recorded the mobility of N-Cam, an unrelated membrane-linked protein that does not accumulate at post-synaptic sites¹⁸ and usually displays diffusive movements on cell surfaces¹⁹. In 8–10-day-old cultures, when synapses are already numerous, N-Cam-bound beads diffuse most of the time (Fig. 1c). Thus, periods of GluR2 immobilization are not due to nonspecific hindrance of bead movement.

We found that zones where confinement of GluR2 occurs are

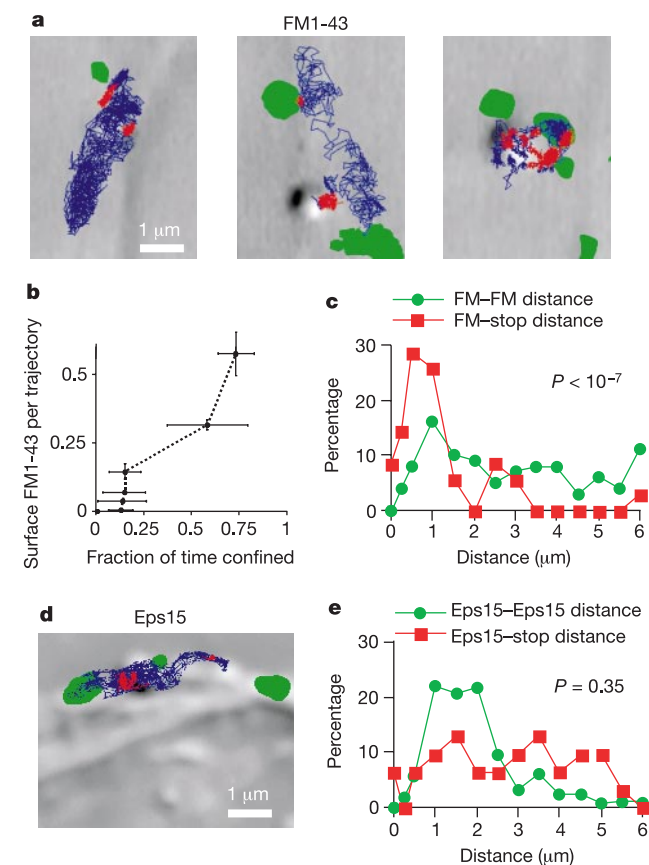


Figure 2 GluR2 stops reversibly at synaptic sites. **a**, Examples of the relation between GluR2 movement and FM1-43 staining (green). Diffusive movement is in blue, confined movement is in red. **b**, Plot of the fraction of surface stained by FM1-43 in the trajectory surface over the fraction of time the trajectory is confined (five recordings per data point). **c**, Normalized distribution of the distances between confinement zones and nearest FM1-43 stain (red squares, 35 observations) and of the distances between nearest-neighbour FM1-43 stains (green circles, 100 observations). Distributions are statistically distinct ($P < 10^{-7}$, Student's *t*-test). **d**, GluR2 movement (colour as in **a**) on a neuron transfected with Eps15-GFP (green). Eps15-GFP stains mark endocytic pits. **e**, Quantification of the distances between confinement zones and nearest Eps15-GFP stain (red squares, 21 observations) and between nearest-neighbour Eps15-GFP stains (green circles, 280 observations).

topographically related to synaptic structures. We located synapses with the pre-synaptic fluorescent marker FM1-43 and simultaneously mapped AMPA receptor movement with a bead (Fig. 2a). Receptors express confined movements more often on neurites that display a high level of FM1-43 staining (Fig. 2b) ($n = 35$, $P < 0.0005$, Spearman's paired test that the proportion of FM1-43 staining and fraction of time anchored are related). For quantification, we measured the distance between sites of confinement and the border of the nearest FM1-43 stain (FM-stop distance) as well as distances between adjacent FM1-43 stains for the concerned neurite (FM-FM distance). The distribution of FM-stop distances was strongly skewed towards shorter distances when compared with the distribution of FM-FM distances (Fig. 2c). On neurites that displayed FM1-43 staining, 51% of the confinement sites ($n = 35$ sites on 24 cells) were less than $0.5\mu\text{m}$ from an FM1-43 stain. Notably, receptors were often seen to halt only briefly in these regions (mean dwell time 24 ± 9 s, $n = 11$) and to switch from one FM region to the other (Fig. 2a). These results show that around synapses there is a region of low receptor diffusion. This may

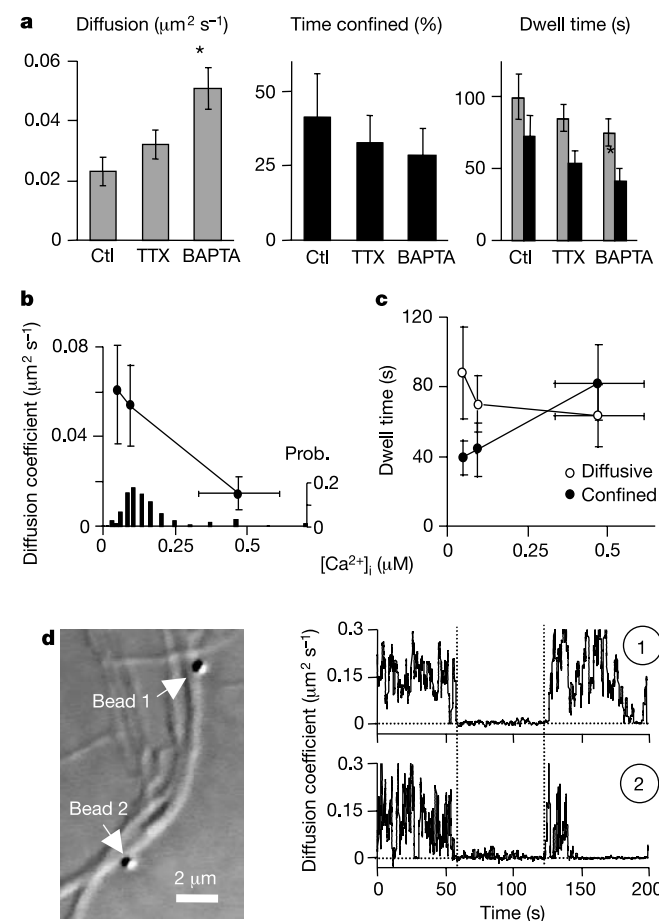


Figure 3 Spontaneous neuronal activity modulates GluR2 mobility. **a**, GluR2 mobility parameters in control conditions (Ctl, 12–17 DIV, $n = 21$), in the presence of $1\mu\text{M}$ TTX (12–15 DIV, $n = 61$), or $1\mu\text{M}$ TTX and intracellular BAPTA (12–13 DIV, $n = 23$). The diffusion coefficient, percentage of time spent in confined state, and dwell time for diffusive (grey bars) and confined (black bars) episodes are shown ($P < 5 \times 10^{-3}$, Student's *t*-test). **b**, Plot of GluR2 diffusion coefficient versus $[\text{Ca}^{2+}]_i$ (14–16 values per data point, 8–12 DIV). Black bars represent the $[\text{Ca}^{2+}]_i$ probability distribution (right axis) for 150 randomly sampled neurites. **c**, Diffusive and confined dwell times versus $[\text{Ca}^{2+}]_i$ for the data in **b**. **d**, Simultaneous recording of two GluR2-bound beads on a single neurite. Left, differential interference contrast image of the neurite (14 DIV). Right, diffusion coefficient versus time for each bead. Dotted lines indicate occasions where both beads stop and resume movement at about the same time.

correspond to reversible AMPA receptor binding to the external border of the post-synaptic density²⁰, in the way we observed for the reversible binding of glycine and mGluR5 (metabotropic glutamate) receptors to extra-synaptic clusters of gephyrin¹⁶ and Homer²¹, respectively.

Confinement zones not associated with FM1-43 staining could correspond to interaction of GluR2 with unstained synapses or with extrasynaptic structures such as receptor clusters^{16,21,22} or sites of endocytosis¹⁰. We tested the latter hypothesis by mapping endocytic pits with EGF-receptor pathway substrate 15 (Eps15) fused to green fluorescent protein (GFP)²³ while measuring GluR2 movement

(Fig. 2d). We found no correlation between Eps15 staining and zones of GluR2 confinement for 18 of 20 beads (Fig. 2e). However, on two occasions receptor immobilization occurred right on top of a pit (not shown). Thus, at best a small fraction of the confined episodes corresponds to arrest in endocytic pits.

Basal neuronal activity in synaptically connected neurons could also control AMPA receptor movement. Tetrodotoxin (TTX), which blocks sodium-dependent action potentials, slightly increased the mean diffusion coefficient and decreased the percentage of time spent in the confined state (Fig. 3a). Buffering intracellular calcium with BAPTA in the presence of TTX further increased diffusion and decreased confinement, mainly by significantly decreasing the mean duration of the confined state. That free intracellular calcium concentration ($[Ca^{2+}]_i$) is involved in the modulation of AMPA receptor mobility was further supported by the finding that receptors move more slowly in neurites having a spontaneously elevated $[Ca^{2+}]_i$ (Fig. 3b, 8–12 DIV). This was mainly due to an increase in the time spent in the confined state at increased $[Ca^{2+}]_i$ (Fig. 3c) ($P < 0.05$, Spearman's test). $[Ca^{2+}]_i$ in neurites distributed mainly (136 of 150 observations) between 0.05 and 0.25 μM (Fig. 3b, black bars). Together, these data indicate that, under resting $[Ca^{2+}]_i$ conditions, AMPA receptor mobility is relatively high, and that $[Ca^{2+}]_i$ has to be significantly raised to slow down receptor movement, as may occur during neuronal activity²⁴. We reasoned that if neuronal activity would modify GluR2 movement through brief elevations of $[Ca^{2+}]_i$, it would do so over the whole neurite in a temporally correlated manner. We recorded simultaneously the movement of two beads several micrometres apart on the same neurite. On a few occasions, we observed temporally correlated transitions from mobile to immobile and to mobile again (Fig. 3d) (on 18 couples of recorded beads, 10 of 110 such transitions occurred within 2 s of each other, whereas only about 2 should have been observed by chance only). These correlated transitions may reflect a spontaneous wave of physiological change in the neurite that freezes all receptors temporarily. Our data suggest that basal neuronal activity and/or $[Ca^{2+}]_i$ regulate, but do not fully determine, movement of AMPA receptors.

Activity-dependent elevation in post-synaptic $[Ca^{2+}]_i$ is known to be a central element in the regulation of synaptic transmission efficacy²⁴. Thus, it participates in both long-term depression (LTD) and potentiation (LTP) of synapses²⁵ as well as in fast exchange of AMPA receptors at synaptic sites²⁶. We investigated the effect of a localized $[Ca^{2+}]_i$ increase on receptor movement. This was achieved by photorelease of a caged Ca^{2+} ionophore with an ultraviolet (UV) laser focused on a small region near the receptor-coupled bead. In 76% of the trials ($n = 47$), activation of the Ca^{2+} ionophore produced a marked and rapid decrease in receptor mobility (Fig. 4a, b, and Supplementary Information). For these trials, the diffusion coefficient decreased on average to 20% (s.d. = 25%, range 0.2–80%) of the pre-stimulus value. The remaining population showed either no significant change in the diffusion coefficient (14% of the trials) or an increase (10% of the trials), as in control experiments ($n = 22$) with no Ca^{2+} ionophore added (Fig. 4b). The decrease in receptor mobility that followed photorelease could be partially prevented if neurons were pre-loaded with the Ca^{2+} chelator BAPTA ($n = 11$) (Fig. 4c), indicating the direct involvement of Ca^{2+} ions. In most cells, the effect lasted throughout the experiment (200 s). A partial recovery was observed in five of the trials. In comparison, the $[Ca^{2+}]_i$ increase induced by the photorelease lasted only a few seconds. The Ca^{2+} -induced receptor immobilization did not depend on the presence of functional synapses, because it was similar before and after synaptogenesis had started in our cultures (Fig. 4c). The Ca^{2+} -induced immobilization was not observed for N-Cam or mGluR5, another glutamate receptor whose basal mobility resembles that of GluR2 (ref. 21).

Activation of mGluR5 with its natural ligand glutamate leads to increased $[Ca^{2+}]_i$ and increased mGluR5 mobility²¹. In contrast,

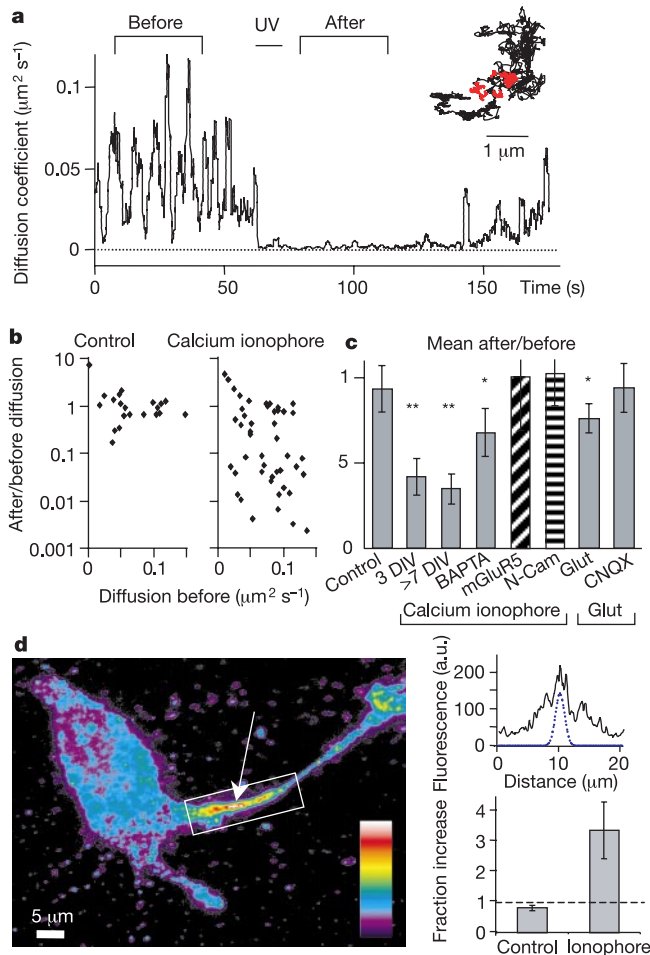


Figure 4 Local rises in intracellular calcium decrease GluR2 mobility and accumulate GluR2. **a**, Diffusion coefficient of GluR2 versus time during the uncaging of a caged calcium ionophore with a train of UV pulses (bar). Inset: GluR2 trajectory before (black) and after (red) uncaging. **b**, Ratios of the diffusion coefficient after and before uncaging versus the latter. (Before and after analysis windows are marked in **a**.) Each dot represents an experiment without (control, vehicle only) or with caged calcium ionophore (2–21 DIV). **c**, Mean values of the ratios of the diffusion coefficient after and before uncaging of caged calcium ionophore or caged glutamate (Glut), as indicated. From left to right: GluR2 with vehicle only (control), GluR2 with caged ionophore at 3 DIV ($n = 21$), at >7 DIV ($n = 26$), pre-loaded with BAPTA, mGluR5 and N-Cam with caged ionophore, GluR2 with caged glutamate without or with CNQX. Double asterisk, $P < 10^{-4}$; asterisk, $P < 0.05$; Student's paired t -test. **d**, Left, pseudo-colour image of the immunostaining of surface GluR2 on a neuron challenged for 5 min with repeated uncaging (arrow) of the caged calcium ionophore. Right, top, linescan intensity profile along the white box indicated in left panel. Dashed line indicates UV radiation intensity profile. a.u., arbitrary units. Right, bottom, mean percentage of increase in fluorescence intensity in the flashed region when compared with an adjacent unflashed region of the same neurite for control experiments (vehicle only, $n = 7$) and with the caged ionophore ($n = 9$) ($P < 0.015$, Student's t -test).

release of caged glutamate in the proximity of GluR2-bound beads induced an average decrease in GluR2 diffusion down to 33% (s.d. = 28%, range 0.3–76%) of the pre-stimulus value in 48% of the trials ($n = 35$) (Fig. 4c). In total, the effect of glutamate was significant, but less robust than that of the caged Ca^{2+} ionophore. The slowdown of receptor movement by glutamate could be prevented by a mixed AMPA/NMDA antagonist (50 μM 6-cyano-7-nitroquinoxaline-2,3-dione, CNQX; $n = 16$) (Fig. 4c), but not by an NMDA-specific antagonist (100 μM D(-)-2-amino-5-phosphonovaleric acid, AP5; $n = 12$) (data not shown).

Finally, we found that a series of repetitive localized $[\text{Ca}^{2+}]_i$ increases triggered by photorelease of the caged ionophore induced within minutes a major local accumulation of GluR2 on the neuronal surface (Fig. 4d). The same protocol did not trigger local accumulation of the coated pit protein Eps15 (not shown), nor was GluR2 accumulation observed when applying UV flashes in the absence of the caged ionophore. The most straightforward explanation for the calcium-induced increase in GluR2 density is that receptors accumulate over time as a result of repetitive calcium-induced immobilization of receptors that pass the zone of uncaging by lateral diffusion. This receptor immobilization on the neuronal surface could arise from calcium-induced increase in the affinity of GluR2 for scaffolding elements²⁷ and/or recruitment of such elements.

GluR2-containing AMPA receptors move rapidly in the plasma membrane of cultured neurons and may cover large distances. They halt regularly and reversibly in confinement zones, mostly corresponding to reversible arrest near synaptic structures, but sometimes also originating from arrest in endocytic pits and from spontaneous neuronal activity. The exchange of AMPA receptors between the post-synaptic density and the extra-synaptic membrane through lateral diffusion could be a prerequisite to the receptors' endo- and exocytotic regulatory processes, which are linked to LTP^{1,2,11,28} and LTD^{3,6,7,9,14} and are likely to occur preferentially outside the synapse. That local Ca^{2+} influx can regulate lateral GluR2 movement and local receptor concentration opens new avenues for the rapid control of AMPA receptor numbers at post-synaptic sites. Regulating the rate at which AMPA receptors can leave or enter synapses could lead to activity-dependent regulation of synaptic transmission. □

Methods

Cell culture and microscopy

Hippocampal neurons from rat embryos at embryonic day 18 were cultured on glass coverslips²⁹. For experiments with mGluR5, cultures were transfected with the complementary DNA encoding Myc-tagged mGluR5 (ref. 21). Video microscopy, latex bead preparation and manipulation with an optical trap were as described^{16,21}. Latex beads (0.5 μm diameter, Polysciences) were coated with anti-GluR2 (Chemicon), anti-N-Cam (a gift of R.-M. Mege) or anti-Myc (Roche diagnostics) antibodies. Neurons were mounted between two coverslips in culture medium with beads and 20 mM HEPES buffer (pH 7.2), kept at 37 °C and visualized under red illumination and differential interference contrast through a 100 $\times$, 1.4 numerical aperture PlanApo objective (Olympus). From each coverslip, we recorded bead trajectories of 200 s each on 1–4 neurons over a period of ~20 min. Bead binding to neurites was virtually suppressed by diluting 100-fold the anti-GluR2 antibody concentration at the bead surface.

Pre-synaptic staining with FM1-43

Neurons (10–17 DIV) were depolarized for 1 min in culture medium supplemented with 40 mM KCl and 10 μM FM1-43 (fluorescent and caged compounds from Molecular Probes), then washed and kept in standard medium containing 1 μM TTX (Sigma). After recording a bead trajectory, FM1-43 staining was visualized in the same recording field by epifluorescence (excitation/emission = 460–480/500–540 nm wavelength) with a Pentamex camera (Princeton Instruments).

Endocytic pit staining with Eps15

Neurons (7–9 DIV) were transfected with the cDNA encoding GFP-tagged Eps15, which is targeted to clathrin-coated pits²³. Bead trajectories were recorded 24 h after transfection. Visualization and quantification procedures were the same as for FM1-43.

Measurements of intracellular free calcium

Neurons were loaded with the dye Fura-2 AM (5 μM) for 20 min at 37 °C. At the start of the experiment, a neurite was chosen and its $[\text{Ca}^{2+}]_i$ determined from standard calibrated

Fura-2 ratiometric imaging (DG4 filter changer, Sutter Instruments). Receptor movement was then recorded from the Fura-2-imaged region.

Flash photolysis

To achieve fast and local rises in intracellular calcium or glutamate levels, we used local photorelease of respectively 1-(4,5-dimethoxy-2-nitrophenyl)ethyl (DMNPE)-caged Ca^{2+} ionophore A-23187 (25 μM) or N-(CNB-caged) L-glutamic acid (25 mM), which were added to the extracellular medium. For photorelease, a pulsed nitrogen UV laser (VSL 337, LSI) was focused as a diffraction-limited spot in the image field. The photorelease protocol consisted of 30 UV flashes (~3- μJ pulses at 1 Hz) close to the tracked bead. For the caged ionophore, GluR2 immobilization was usually achieved with a single UV flash. Control $[\text{Ca}^{2+}]_i$ imaging experiments performed with Fluo-3 in the presence of the caged ionophore verified that a single UV flash produced a rapid, substantial (>0.5 μM) and transient (mean decay time constant 2.5 s) increase in dendritic $[\text{Ca}^{2+}]_i$. For quantification, in each recording we compared the mean diffusion coefficient during 30-s periods before and after UV illumination. For BAPTA loading, cells were incubated for 15 min with 5 μM BAPTA-AM (Molecular Probes). For GluR2 accumulation, during 5 min, one burst (50 Hz) of ten UV pulses was delivered every 30 s. Cells were fixed immediately afterwards with 4% paraformaldehyde-sucrose to avoid permeabilization, and processed for immunostaining with anti-GluR2 and secondary antibodies. Experiments were performed on neurons at 4–6 DIV that were transfected with the cDNA encoding a GFP-tagged GluR2, to help cell recognition.

Bead tracking and data analysis

Positions of GluR2-bound beads were tracked on digitized video images^{16,21}. Diffusion coefficients for whole or parts of trajectories were calculated by fitting the first three points of the mean squared displacement (MSD) curves versus time. Plots of diffusion coefficients versus time were derived from MSDs calculated on contiguous trajectory stretches of 50 points. Confined movement is a period during which receptor movement is contained within a region smaller than that expected from free brownian movement. To detect periods of confined receptor movement, we calculated a confinement index¹⁵ as previously described^{16,21}. Confined periods were detected as stretches of positions having an index greater than 3.16 for 2.7 s or more. This threshold reflects a 99.3% likelihood to pertain to a confined movement. Distances from confinement zone to the nearest border of an FM1-43 stain were calculated from FM1-43 staining images made binary and overlaid with the corresponding bead trajectory with Metamorph 4.0 software (Universal Imaging Corporation). Data are given as mean $\pm$ s.e.m., unless noted.

Received 15 February; accepted 2 April 2002; doi:10.1038/nature00780.

- Hayashi, Y. *et al.* Driving AMPA receptors into synapses by LTP and CaMKII: requirement for GluR1 and PDZ domain interaction. *Science* **287**, 2262–2267 (2000).
- Shi, S. H. *et al.* Rapid spine delivery and redistribution of AMPA receptors after synaptic NMDA receptor activation. *Science* **284**, 1811–1816 (1999).
- Carroll, R. C., Lissin, D. V., von Zastrow, M., Nicoll, R. A. & Malenka, R. C. Rapid redistribution of glutamate receptors contributes to long-term depression in hippocampal cultures. *Nature Neurosci.* **2**, 454–460 (1999).
- Shi, S., Hayashi, Y., Esteban, J. A. & Malinow, R. Subunit-specific rules governing AMPA receptor trafficking to synapses in hippocampal pyramidal neurons. *Cell* **105**, 331–343 (2001).
- Carroll, R. C., Beattie, E. C., von Zastrow, M. & Malenka, R. C. Role of AMPA receptor endocytosis in synaptic plasticity. *Nature Rev. Neurosci.* **2**, 315–324 (2001).
- Luthi, A. *et al.* Hippocampal LTD expression involves a pool of AMPARs regulated by the NSF–GluR2 interaction. *Neuron* **24**, 389–399 (1999).
- Beattie, E. C. *et al.* Regulation of AMPA receptor endocytosis by a signalling mechanism shared with LTD. *Nature Neurosci.* **3**, 1291–1300 (2000).
- Lin, J. W. *et al.* Distinct molecular mechanisms and divergent endocytotic pathways of AMPA receptor internalization. *Nature Neurosci.* **3**, 1282–1290 (2000).
- Wang, Y. T. & Linden, D. J. Expression of cerebellar long-term depression requires postsynaptic clathrin-mediated endocytosis. *Neuron* **25**, 635–647 (2000).
- Man, Y. H. *et al.* Regulation of AMPA receptor-mediated synaptic transmission by clathrin-dependent receptor internalization. *Neuron* **25**, 649–662 (2000).
- Lu, W. *et al.* Activation of synaptic NMDA receptors induces membrane insertion of new AMPA receptors and LTP in cultured hippocampal neurons. *Neuron* **29**, 243–254 (2001).
- Sheng, M. & Lee, S. H. AMPA receptor trafficking and the control of synaptic transmission. *Cell* **105**, 825–828 (2001).
- Noel, J. *et al.* Surface expression of AMPA receptors in hippocampal neurons is regulated by an NSF-dependent mechanism. *Neuron* **23**, 365–376 (1999).
- Luscher, C. *et al.* Role of AMPA receptor cycling in synaptic transmission and plasticity. *Neuron* **24**, 649–658 (1999).
- Simson, R., Sheets, E. D. & Jacobson, K. Detection of temporary lateral confinement of membrane proteins using single-particle tracking analysis. *Biophys. J.* **69**, 989–993 (1995).
- Meier, J., Vannier, C., Sergé, A., Triller, A. & Choquet, D. Fast and reversible trapping of surface glycine receptors by gephyrin. *Nature Neurosci.* **4**, 253–260 (2001).
- Fletcher, T. L., De Camilli, P. & Banker, G. Synaptogenesis in hippocampal cultures: evidence indicating that axons and dendrites become competent to form synapses at different stages of neuronal development. *J. Neurosci.* **14**, 6695–6706 (1994).
- Kiss, J. Z. & Muller, D. Contribution of the neural cell adhesion molecule to neuronal and synaptic plasticity. *Rev. Neurosci.* **12**, 297–310 (2001).
- Simson, R. *et al.* Structural mosaicism on the submicron scale in the plasma membrane. *Biophys. J.* **74**, 297–308 (1998).
- Nusser, Z. AMPA and NMDA receptors: similarities and differences in their synaptic distribution. *Curr. Opin. Neurobiol.* **10**, 337–341 (2000).
- Sergé, A., Fourgeaud, L., Hémar, A. & Choquet, D. Receptor activation and homer differentially control the lateral mobility of mGluR5 in the neuronal membrane. *J. Neurosci.* (in the press).

22. Pickard, L., Noel, J., Henley, J. M., Collingridge, G. L. & Molnar, E. Developmental changes in synaptic AMPA and NMDA receptor distribution and AMPA receptor subunit composition in living hippocampal neurons. *J. Neurosci.* **20**, 7922–7931 (2000).
23. Benmerah, A., Poupon, V., Cerf-Bensussan, N. & Dautry-Varsat, A. Mapping of Eps15 domains involved in its targeting to clathrin-coated pits. *J. Biol. Chem.* **275**, 3288–3295 (2000).
24. Ghosh, A. & Greenberg, M. E. Calcium signalling in neurons: molecular mechanisms and cellular consequences. *Science* **268**, 239–247 (1995).
25. Yang, S. N., Tang, Y. G. & Zucker, R. S. Selective induction of LTP and LTD by postsynaptic $[Ca^{2+}]_i$ elevation. *J. Neurophysiol.* **81**, 781–787 (1999).
26. Liu, S. Q. & Cull-Candy, S. G. Synaptic activity at calcium-permeable AMPA receptors induces a switch in receptor subtype. *Nature* **405**, 454–458 (2000).
27. Lee, H. K., Barbarosie, M., Kameyama, K., Bear, M. F. & Huganir, R. L. Regulation of distinct AMPA receptor phosphorylation sites during bidirectional synaptic plasticity. *Nature* **405**, 955–959 (2000).
28. Malinow, R., Mainen, Z. F. & Hayashi, Y. LTP mechanisms: from silence to four-lane traffic. *Curr. Opin. Neurobiol.* **10**, 352–357 (2000).
29. Hémar, A., Olivo, J. C., Williamson, E., Saffrich, R. & Dotti, C. G. Dendroaxonal transcytosis of transferrin in cultured hippocampal and sympathetic neurons. *J. Neurosci.* **17**, 9026–9034 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

This work was supported by grants from the Centre National de la Recherche Scientifique and the Conseil Régional d'Aquitaine. A.J.B. was supported by the Fondation pour la Recherche Médicale and an EC (European Commission) Marie Curie Training fellowship. We thank P. Osten for his gift of GluR2–GFP cDNA, A. Benmerah for the gift of the Esp15–GFP cDNA, and R.-M. Mège for the gift of anti-N-Cam. We thank P. Ascher for his support during early phases of this work; C. Mulle, A. Hémar and L. Cognet for their comments on the manuscript; and F. Rossignol for cultures of hippocampal neurons.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.C. (e-mail: dchoquet@u-bordeaux2.fr).

Annexin II light chain regulates sensory neuron-specific sodium channel expression

Kenji Okuse*, Misbah Malik-Hall*, Mark D. Baker*, W-Y. Louisa Poon*, Haeyoung Kong†, Moses V. Chao† & John N. Wood*

* Department of Biology, University College London, Gower Street, London WC1E 6BT, UK

† Molecular Neurobiology Program, Skirball Institute for Biomolecular Medicine, Departments of Cell Biology and Physiology and Neuroscience, New York University School of Medicine, New York, New York 10016, USA

The tetrodotoxin-resistant sodium channel $Na_v1.8$ /SNS is expressed exclusively in sensory neurons and appears to have an important role in pain pathways^{1,2}. Unlike other sodium channels, $Na_v1.8$ is poorly expressed in cell lines even in the presence of accessory β -subunits³. Here we identify annexin II light chain^{4,5} (p11) as a regulatory factor that facilitates the expression of $Na_v1.8$. p11 binds directly to the amino terminus of $Na_v1.8$ and promotes the translocation of $Na_v1.8$ to the plasma membrane, producing functional channels. The endogenous $Na_v1.8$ current in sensory neurons is inhibited by antisense downregulation of p11 expression. Because direct association with p11 is required for functional expression of $Na_v1.8$, disrupting this interaction may be a useful new approach to down-regulating $Na_v1.8$ and effecting analgesia⁶.

Voltage-gated sodium channels initiate and propagate action potentials in excitable cells. Ten distinct pore-forming α -subunits of voltage-gated sodium channels have been identified⁷. Expression of the tetrodotoxin-resistant (TTX-resistant) sodium channel

$Na_v1.8$ (also known as SNS/PN3 or SCN10a) is restricted to small-diameter C-fibre-associated sensory neurons¹. Behavioural studies on $Na_v1.8$ null mutant mice have shown that $Na_v1.8$ helps in the detection of noxious thermal, mechanical and inflammatory stimuli².

The α -subunits of sodium channels are known to associate with auxiliary β -subunits that promote functional channel expression^{3,8}. Microinjection of $Na_v1.8$ complementary DNA into the nuclei of superior cervical ganglion (SCG) neurons results in the robust expression of a sodium current that shows exactly the same biophysical properties as those observed in dorsal root ganglia (DRG) neurons⁹. In contrast, Chinese hamster ovary (CHO), African green monkey kidney (COS-7), human embryonic kidney (HEK-293) and other mammalian cell lines express few or no detectable $Na_v1.8$ channels after introduction of $Na_v1.8$ cDNA⁹, and the expressed channel shows different properties from the endogenous DRG current in these cell types¹⁰ even in the presence of auxiliary β -subunits (M.D.B., unpublished observations). These observations suggest that $Na_v1.8$ requires a distinct specific subunit or permissive factor different from known β -subunits to promote $Na_v1.8$ functional expression on the plasma membrane.

Here we report the use of the yeast two-hybrid interactive screen to identify p11 (the annexin II light chain)^{4,5} as a previously unknown regulatory factor for $Na_v1.8$ functional expression. A rat sensory neuron cDNA library¹¹ was used to screen for proteins that interact with the N-terminal intracellular domain of rat $Na_v1.8$. Five identical positive clones coding for a full-length p11 were identified through their interaction with the N terminus of $Na_v1.8$. To test whether p11 binds to the N-terminal intracellular domain of $Na_v1.8$ *in vitro*, we expressed the green fluorescent protein (GFP)–p11 fusion protein in COS-7 cells, and expressed the N-terminal domain of $Na_v1.8$ as a glutathione S-transferase (GST) fusion protein, GST–SNS(I). GST–SNS(I) pulled down the GFP–p11 fusion proteins specifically and efficiently in an GST pull-down assay¹² (see Supplementary Information).

We examined the tissue distribution of the p11 transcript. North-

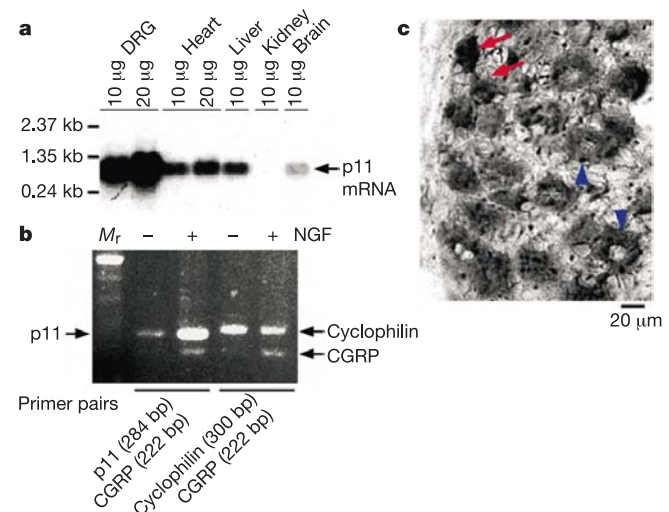


Figure 1 p11 mRNA is expressed in DRG neurons and is upregulated by NGF. **a**, Total RNA isolated from 2-week-old rat DRG, heart, liver, kidney and whole brain were subjected to Northern blot analysis. High levels of p11 mRNA are present in DRG and other tissues. **b**, RT–PCR analysis of p11 mRNA extracted from DRG neurons grown in the absence (–) or presence (+) of NGF for 7 days. A large increase of p11 mRNA is observed in response to NGF treatment. **c**, *In situ* hybridization shows that p11 mRNA is expressed in small- and large-diameter neurons in DRG. Damage-sensing small-diameter neurons (arrows in red) expressed p11 mRNA as well as large diameter neurons (arrowheads in blue). M_r , relative molecular mass.

ern blot analysis showed high levels of expression of p11 messenger RNA in DRG, modest expression in heart and liver, and weak expression in brain isolated from 2-week-old rats (Fig. 1a). Polymerase chain reaction after reverse transcription of RNA (RT-PCR) showed a large increase of p11 mRNA in cultured rat DRG neurons treated with nerve growth factor (NGF), which is known to cause decreases in thermal, chemical and mechanical thresholds of pain perception in animal models^{13,14} (Fig. 1b). *In situ* hybridization was performed on a section of 2-week-old rat DRG. An antisense p11 probe demonstrated strong staining in both small and large

diameter neurons (Fig. 1c). Combined immunohistochemistry with anti-Nav1.8 polyclonal antibody SNS11 (ref. 15) showed that most (>98%) of the Nav1.8-positive cells also expressed p11 mRNA.

To examine the effect of p11 on Nav1.8 trafficking, we transfected CHO-SNS22 (a cell line stably transfected with rat Nav1.8 cDNA that expresses Nav1.8 mRNA and cytosolic immunoreactive protein (Fig. 2a, arrows), but does not generate Nav1.8 current) with a GFP-p11 fusion cDNA. The expression of p11 protein could thus be detected as a GFP signal. The green fluorescence was localized specifically in the plasma membrane (Fig. 2b, arrowhead). In the same cell, Nav1.8-like immunoreactivity (red fluorescence) translocated to the plasma membrane (Fig. 2a, arrowhead). The merged picture shows co-expression of p11 and Nav1.8 in the plasma membrane as indicated by a yellow colour (Fig. 2c, arrowhead). Densitometric analysis of Nav1.8-like immunoreactivity of the GFP-p11 fusion or GFP-protein-expressed CHO-SNS22 cells showed that 16.5% (s.e.m., 1.2; $n = 30$) of Nav1.8-like immunoreactivity moved to the plasma membrane fraction after the expression of GFP-p11 fusion protein (Fig. 2d), but only 4.3% (s.e.m., 0.4; $n = 30$) of Nav1.8-like immunoreactivity localized on the plasma membrane in the GFP-expressing CHO-SNS22 cells (Fig. 2e). These data demonstrate that p11 promotes the translocation of Nav1.8 protein to the extracellular membrane.

In nine of 42 CHO-SNS22 cells transfected with GFP-p11 cDNA expression vector, TTX-resistant currents were found that very closely resembled the Nav1.8 Na^+ current recorded from transfected COS-7 cells¹⁰, both in terms of voltage-dependence and kinetics. An example of a family of TTX-resistant inward currents is shown in Fig. 3a. The currents began to activate around -5 mV, and peaked at $+40$ mV (Fig. 3b). Although no TTX-resistant inward

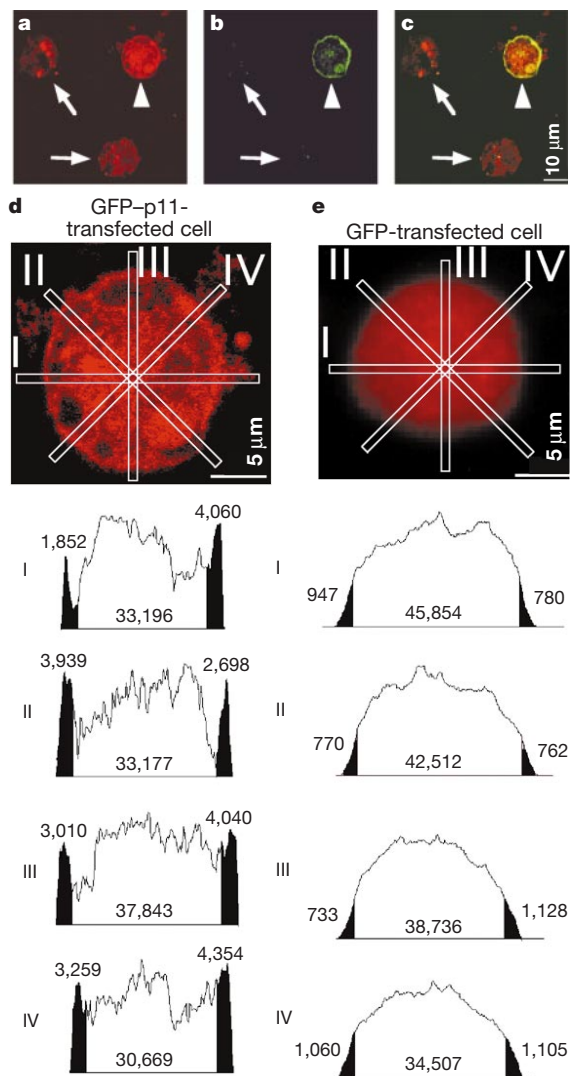


Figure 2 p11 regulates trafficking of Nav1.8 from cytosol to plasma membrane. **a**, Nav1.8-like immunoreactivity in CHO-SNS22 cells visualized by anti-Nav1.8 polyclonal antibody SNS11 and anti-rabbit IgG rhodamine. **b**, transient expression of GFP-p11 fusion protein in CHO-SNS22 cells transfected with GFP-p11 cDNA expression vector by lipofection. **c**, Merged picture of **a** and **b** showing co-expression of p11 and Nav1.8 in the plasma membrane. **d**, **e**, Microscopic images were quantitated using the public-domain NIH Image program (<http://rsb.info.nih.gov/ni-image/>). Typical pictures of Nav1.8-like immunoreactivity obtained from GFP-p11 fusion protein or GFP control protein expressed CHO-SNS22 cells are shown above. Densitometric measurements were taken along four axes, each 45° apart, through the whole cross-section of the cell. For calculation purposes, the thickness of the membrane was set 1.5 μm from the edge of the fluorescence. Percentage translocation to the cytosol was calculated by dividing the number of grains associated with the membrane by the total number of grains in the cell, as indicated, and averaging the results from four axes using 30 cells.

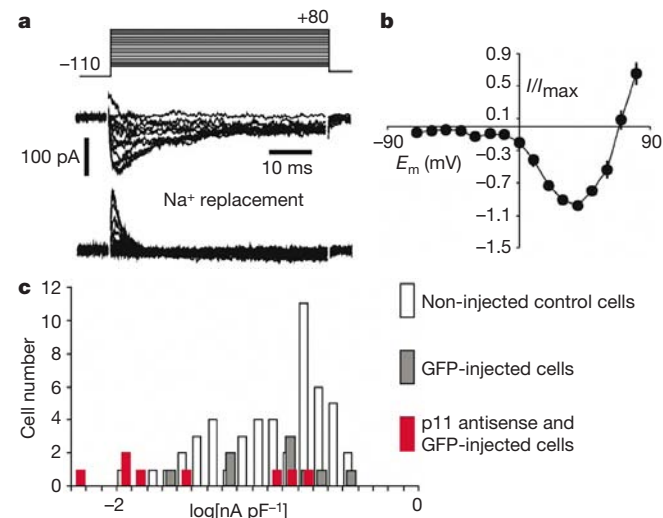


Figure 3 p11 is required for expression of TTX-resistant inward currents in CHO-SNS22 and DRG neurons. **a**, High-threshold TTX-resistant Na^+ current recorded from fluorescent CHO-SNS22 cells after transfection with GFP-p11 cDNA expression vector. Na^+ current has characteristically slow kinetics, and inward current is abolished by removing extracellular Na^+ ions. Pulse protocol is shown above. **b**, Average current (I/I_{max})–voltage (E_m) relationship for the Na^+ current in CHO-SNS22 cells ($n = 5$). The threshold for activation is close to -5 mV, and the current peaks at $+40$ mV. **c**, p11 antisense mRNA expression in DRG neurons caused a loss of Nav1.8 current density. The histogram shows cell number against log[current density] for control and cDNA-injected neurons. Two-tailed, unpaired t -test for the log[current density] of neurons injected with GFP expression vector only and p11 antisense mRNA expression vector showed significant reduction in Nav1.8 current ($P < 0.02$, Student's two-tailed t -test).

currents have ever been previously recorded from either non-transfected ($n = 41$) or GFP-transfected ($n = 40$) CHO-SNS22 cells, after p11 transfection a little over 20% of the cells generated small Na^+ currents. This is a highly statistically significant finding ($P < 0.002$ versus GFP-transfected or control non-transfected cells, Fisher exact test).

To test the possible regulatory role of p11 on $\text{Na}_v1.8$ channels in sensory neurons, we microinjected the p11 antisense expression vector, pBS-GFP/AS(p11), into the nuclei of DRG neurons in culture. Immunohistochemistry using anti-p11 polyclonal antibodies confirmed an efficient reduction of p11-like immunoreactivity in DRG neurons by the introduction of pBS-GFP/AS(p11) (see Supplementary Information). The introduction of pBS-GFP/AS(p11) also caused a great loss of $\text{Na}_v1.8$ current (Fig. 3c). The mean peak Na^+ current density was reduced in pBS-GFP/AS(p11) injected neurons ($63.1 \pm 24.5 \text{ pA pF}^{-1}$, mean $\pm$ s.e.m.; $n = 8$) when compared with control neurons injected with GFP only ($179.2 \pm 40.3 \text{ pA pF}^{-1}$; $n = 9$, $P < 0.04$; Student's two-tailed t -test). In contrast, the residual maximum K^+ current density derived from currents recorded on stepping to $+80 \text{ mV}$ were not significantly affected by pBS-GFP/AS(p11) ($61.7 \pm 17.6 \text{ pA pF}^{-1}$ in control cells versus $44.8 \pm 5.5 \text{ pA pF}^{-1}$ in injected cells (mean $\pm$ s.e.m.; $P = 0.4$; Student's two-tailed t -test), suggesting that the effect of antisense was specific. We also examined the effect of pBS-GFP/AS/p11 on TTX-sensitive currents in ND7/23 cells. TTX-sensitive current densities were unaffected (control, $-1.27 \pm 0.12 \log[\text{nA pF}^{-1}]$; antisense-treated, $-1.52 \pm 0.12 \log[\text{nA pF}^{-1}]$; $n = 12$, $P > 0.1$; Student's unpaired two-tailed t -test), suggesting that p11 is not required for the expression of other TTX-sensitive sodium-channel subtypes.

Annexin II light chain protein p11 thus has an important role in promoting expression of the sensory-neuron-specific voltage-gated sodium channel $\text{Na}_v1.8$. p11 is present in a variety of cells separately or as a heterotetramer composed of two copies of the annexin II heavy chain, also known as calpactin I, lipocortin II, or p36 (refs 4, 5). Annexins have been suggested to have inherent channel activity and have also been implicated in regulating the expression of other channels. Annexin VII has intrinsic calcium channel activity¹⁶, whereas annexin II has been shown to alter the activity of osmotically-regulated chloride channels in endothelial cells¹⁷.

Increased functional expression of $\text{Na}_v1.8$ is associated with diminished pain thresholds¹⁸, so the actions of inflammatory mediators on p11 expression are of interest. In non-neuronal epithelial cells, nitric oxide is known to induce the expression of p11 (ref. 19) as well as having an important role in the development of peripheral hyperalgesia²⁰. Glucocorticoids are also known to regulate expression of members of the annexin family. In human epithelial cells, p11 is induced by the application of dexamethasone²¹. It is therefore possible that glucocorticoids induced as a response to inflammation may also upregulate $\text{Na}_v1.8$ expression. Studies using PC12 cells have found that dexamethasone upregulates voltage-gated sodium channel activity²². Transforming growth factor α (TGF- α) has also been shown to stimulate expression of p11 in an epithelial cell line²³. Treatment with NGF caused upregulation of p11 mRNA level in cultured DRG neurons in our study. p11 is already known to be induced by NGF in PC12 cells²⁴ and NGF is also known to increase sodium current density in both PC12 cells²² and DRG neurons²⁵. There is a small increase in the amount of membrane-associated $\text{Na}_v1.8$ protein in DRG on NGF treatment, whereas the level of $\text{Na}_v1.8$ mRNA appears to be unchanged²⁶. Taken together, these observations suggest that p11 may increase translocation of $\text{Na}_v1.8$ into the plasma membrane in response to tissue damage via growth factors such as NGF and TGF- α , or inflammatory regulators such as glucocorticoid hormones, and the increased $\text{Na}_v1.8$ activity may contribute to the hyperexcitability of sensory neurons. Upregulation of $\text{Na}_v1.8$ activity has been suggested as a mechanism for lowering peripheral pain thresholds in

inflammatory pain states²⁷. The present data suggest that disrupting p11- $\text{Na}_v1.8$ interactions may provide a new way of lowering the expression of TTX-resistant sodium channels in nociceptive neurons, and thus producing analgesia. □

Methods

Yeast two-hybrid screen

A two-hybrid interaction screen was performed with the N-terminal intracellular domain (amino-acid position 1–127) of $\text{Na}_v1.8$ and a rat P1 DRG cDNA library as described¹¹. Approximately 5×10^6 yeast transformants were screened and five identical positive clones coding for a full length p11 were identified. The clone included a 51-base-pair (bp) 5'-UTR, a 288-bp coding region, and a 450-bp 3'-UTR of the rat p11 gene. To verify that p11 interacts specifically with the N-terminal intracellular domain of $\text{Na}_v1.8$, the rescued p11-coding plasmid DNA was re-introduced into other strains of yeast containing different intracellular domains of $\text{Na}_v1.8$ as baits. Direct interaction between p11 and N-terminal domain of $\text{Na}_v1.8$ *in vitro* was assessed using GST pull-down assays¹². Full details of experimental methods are given in the Supplementary Information.

Immunofluorescence analysis

A stably transformed CHO cell line (CHO-SNS22 cells) that expresses rat $\text{Na}_v1.8$ protein in the cytosol was transfected with the expression plasmid pBS-GFP/p11 by lipofection. Three days after transfection, the cells were fixed with 4% paraformaldehyde for 15 min on ice and subsequently incubated with anti- $\text{Na}_v1.8$ polyclonal antibody (SNS11). The cells were washed with phosphate-buffered saline (PBS) and incubated with rhodamine-labelled anti-rabbit IgG before analysis with a confocal microscope.

In situ hybridization

A 284-bp p11 PCR fragment was subcloned into pGEM-T Easy (Promega), and DIG-UTP labelled sense or antisense cRNA probe were generated using T7 RNA polymerase for *in situ* hybridization studies.

NGF regulation of p11

DRG neurons from 2-week-old rats were cultured with NGF (50 ng ml^{-1}) or grown in the absence of NGF and in the presence of rabbit anti-NGF antiserum. Total RNA extracted from the culture was treated with DNase I and cDNA was synthesized with Superscript using random hexamer primers. PCR was performed with the primer pair specific for p11 (284 bp): 5'-CATCCCAATGGAGCATGC-3' and 5'-CTACTTCTTCGCTTCATGTG TACTAC-3'.

Electrophysiology

Membrane currents were recorded from CHO-SNS22 cells or cultured DRG neurons using the whole-cell patch-clamp technique. CHO-SNS22 cells and DRG neurons were held at -90 mV . Voltage-clamp protocols incorporated a negative pre-pulse to -110 mV , and the cell was subsequently stepped to more depolarized potentials for 50 ms (up to a final value of $+80 \text{ mV}$), in 10-mV increments. All experiments were performed at room temperature. ($20\text{--}25^\circ\text{C}$)

Antisense studies for p11 in cultured DRG neurons

The 309-bp *NcoI* fragment of p11 was cloned in the 3' to 5' direction into *NcoI* restriction site in pBS500 vector, resulting in an expression system for a sense-GFP/antisense-p11 fusion RNA, pBS-GFP/AS(p11). That was injected into nuclei of 2-week-old rat DRG small-diameter neurons using an Eppendorf microinjector.

Received 29 January; accepted 28 March 2002; doi:10.1038/nature00781.

1. Akopian, A. N., Sivillotti, L. & Wood, J. N. A tetrodotoxin-resistant voltage-gated sodium channel expressed by sensory neurons. *Nature* **379**, 257–262 (1996).
2. Akopian, A. N. *et al.* The tetrodotoxin-resistant sodium channel SNS plays a specialized role in pain pathways. *Nature Neurosci.* **2**, 541–548 (1999).
3. Isom, L. L. Sodium channel beta subunits: anything but auxiliary. *Neuroscientist* **7**, 42–54 (2000).
4. Waisman, D. M. Annexin II tetramer: structure and function. *Mol. Cell. Biochem.* **149/150**, 301–322 (1995).
5. Osborn, M., Johnsson, N., Wehland, J. & Weber, K. The submembranous location of p11 and its interaction with the p36 substrate of pp60 src kinase *in situ*. *Exp. Cell. Res.* **175**, 81–96 (1988).
6. Waxman, S. G. *et al.* Sodium channels, excitability of primary sensory neurons, and the molecular basis of pain. *Muscle Nerve* **22**, 1177–1187 (1999).
7. Goldin, A. L. *et al.* Nomenclature of voltage-gated sodium channels. *Neuron* **28**, 365–368 (2000).
8. Isom, L. L. *et al.* Functional co-expression of the beta 1 and type IIA alpha subunits of sodium channels in a mammalian cell line. *J. Biol. Chem.* **270**, 3306–3312 (1995).
9. England, S., Okuse, K., Ogata, N. & Wood, J. N. Heterologous expression of the sensory neuron-specific sodium channel (SNS) α -subunit in rat sympathetic neurones. *J. Physiol.* **511**, 124P (1998).
10. Fitzgerald, E. M. *et al.* cAMP-dependent phosphorylation of the tetrodotoxin-resistant voltage-dependent sodium channel SNS. *J. Physiol.* **516**, 433–446 (1999).
11. Kong, H. *et al.* An evolutionarily conserved transmembrane protein that is a novel downstream target of neurotrophin and ephrin receptors. *J. Neurosci.* **21**, 176–185 (2001).
12. Liu, C. J., Dib-Hajj, S. D. & Waxman, S. G. Fibroblast growth factor homologous factor 1B binds to the C terminus of the tetrodotoxin-resistant sodium channel rNav1.9a (NaN). *J. Biol. Chem.* **276**, 18925–18933 (2001).
13. Lewin, G. R., Rueff, A. & Mendell, L. M. Peripheral and central mechanisms of NGF-induced hyperalgesia. *Eur. J. Neurosci.* **6**, 1903–1912 (1994).
14. Black, J. A. *et al.* NGF has opposing effects on Na^+ channel III and SNS gene expression in spinal sensory neurons. *Neuroreport* **8**, 2331–2335 (1997).

15. Fang, X. *et al.* Sensory and electrophysiological properties of DRG neurones with SNS-like immunoreactivity (SNS-LI) in rats. *Soc. Neurosci. Abstr.* no. 819.5 (2001).
16. Pollard, H. B. & Rojas, E. Calcium activated synexin forms highly selective voltage-gated calcium channels in phosphatidylserine bilayer membranes. *Proc. Natl Acad. Sci. USA* **85**, 2974–2978 (1988).
17. Nilius, B. *et al.* Annexin II modulates volume activated chloride channels in vascular endothelial cells. *J. Biol. Chem.* **271**, 30631–30636 (1996).
18. Baker, M. D. & Wood, J. N. Involvement of Na⁺ channels in pain pathways. *Trends Pharmacol. Sci.* **22**, 27–31 (2001).
19. Pawliczak, R. *et al.* p11 expression in human bronchial epithelial cells is increased by nitric oxide in a cGMP-dependent pathway involving protein kinase G activation. *J. Biol. Chem.* **276**, 44613–44620 (2001).
20. Aley, K. O., McCarter, G. & Levine, J. D. Nitric oxide signalling in pain and nociceptor sensitization in the rat. *J. Neurosci.* **18**, 7008–7014 (1998).
21. Yao, X. L. *et al.* Dexamethasone alters arachidonate release from human epithelial cells by induction of p11 protein synthesis and inhibition of phospholipase A2 activity. *J. Biol. Chem.* **274**, 17202–17208 (1999).
22. Garber, S. S., Hoshi, T. & Aldrich, R. W. Regulation of ionic currents in pheochromocytoma cells by nerve growth factor and dexamethasone. *J. Neurosci.* **9**, 3976–3987 (1989).
23. Akiba, S. *et al.* Transforming growth factor- α stimulates prostaglandin generation through cytosolic phospholipase A2 under the control of p11 in rat gastric epithelial cells. *Br. J. Pharmacol.* **131**, 1004–1010 (2000).
24. Masiakowski, P. & Shooter, E. M. Nerve growth factor induces the genes for two proteins related to a family of calcium-binding proteins in PC12 cells. *Proc. Natl Acad. Sci. USA* **85**, 1277–1281 (1988).
25. Fjell, J. *et al.* In vivo NGF deprivation reduces SNS expression and TTX-R sodium currents in IB4-negative DRG neurons. *J. Neurophysiol.* **81**, 803–810 (1999).
26. Okuse, K. *et al.* Regulation of expression of the sensory neuron-specific sodium channel SNS in inflammatory and neuropathic pain. *Mol. Cell. Neurosci.* **10**, 196–207 (1997).
27. Gold, M. S. Tetrodotoxin-resistant Na⁺ currents and inflammatory hyperalgesia. *Proc. Natl Acad. Sci. USA* **96**, 7645–7649 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank the MRC, the Wellcome Trust and the NIH (grants supporting M.C. and H.K.) for their generous support. We thank Stephen E. Moss and A. Schmidt for helpful advice.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.N.W. (e-mail: J.Wood@ucl.ac.uk).

Filamentous phage integration requires the host recombinases XerC and XerD

Kathryn E. Huber & Matthew K. Waldor

Division of Geographic Medicine/Infectious Diseases, New England Medical Center and Department of Microbiology, Tufts University School of Medicine and Howard Hughes Medical Institute, 750 Washington Street, Boston, Massachusetts 02111, USA

Many bacteriophages and animal viruses integrate their genomes into the chromosomal DNA of their hosts as a method of promoting vertical transmission. Phages that integrate in a site-specific fashion encode an integrase enzyme that catalyses recombination between the phage and host genomes^{1,2}. CTX ϕ is a filamentous bacteriophage that contains the genes encoding cholera toxin, the principal virulence factor of the diarrhoea-causing Gram-negative bacterium *Vibrio cholerae*³. CTX ϕ integrates into the *V. cholerae* genome in a site-specific manner^{4,5}; however, the ~6.9-kilobase (kb) CTX ϕ genome does not encode any protein with significant homology to known recombinases. Here we report that XerC and XerD, two chromosome-encoded recombinases that ordinarily function to resolve chromosome dimers at the *dif* recombination site⁶, are essential for CTX ϕ

integration into the *V. cholerae* genome. The CTX ϕ integration site was found to overlap with the *dif* site of the larger of the two *V. cholerae* chromosomes. Examination of sequences of the integration sites of other filamentous phages indicates that the XerCD recombinases also mediate the integration of these phage genomes at *dif*-like sites in various bacterial species.

Recombination between CTX ϕ DNA and the *V. cholerae* chromosome was studied by introducing a kanamycin-resistance marked plasmid form of CTX ϕ , pCTX-Kn, into strain 2740-80, an El Tor *V. cholerae* isolate that lacks the CTX prophage⁴. In accord with previous studies^{4,5}, pCTX-Kn integrated into a specific site on the larger of the two *V. cholerae* chromosomes (chrI) (Fig. 1). The DNA sequences of the right and left prophage–chromosome junctions, designated *attL* and *attR* in Fig. 1, were determined. Recombination was found to occur between 15-base-pair (bp) sequences (Fig. 1) of near identity that lie within the phage attachment site (*attP*) and the chromosome attachment site (*attB*). This recombination reaction seemed to be conservative because the recombination products showed no evidence of either addition or loss of sequence^{1,2}.

Because the CTX ϕ genome lacks any gene that would encode a recombinase, we proposed that CTX ϕ integration might depend on a chromosome-encoded enzyme. The fact that the CTX prophage is found about 180° from the putative origin of replication on chrI in El Tor *V. cholerae* isolates⁷, and also that the *attB* sequence is similar to that of *Escherichia coli dif* (see Fig. 4), led us to examine whether *V. cholerae* XerC and XerD recombinases might mediate CTX ϕ integration. In *E. coli*, tyrosine recombinases XerC and XerD act at the 28-bp chromosomal *dif* site, located opposite the origin of replication, to generate monomers from dimeric circular chromosomes commonly formed by homologous recombination^{6,8}. The *V. cholerae* *xerC* and *xerD* orthologues⁷ encode proteins that are 53% and 68% identical in amino acid sequence to the *E. coli* XerC and XerD proteins, respectively. We found that both *xerC* and *xerD* were required for the integration of CTX ϕ into the *V. cholerae* genome. When pCTX-Kn was introduced by electroporation into derivatives of strain 2740-80 containing insertion mutations in either *xerC* or *xerD*, independent Kn^R transformants contained only pCTX-Kn

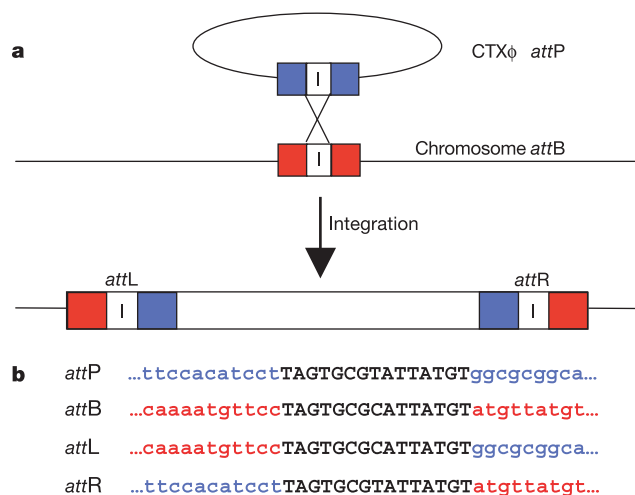


Figure 1 Site-specific integration of CTX ϕ DNA into chrI of strain 2740-80. **a**, Schematic depiction of the recombination between CTX ϕ and chromosomal DNA. The phage attachment site, *attP* (in blue), and the chromosomal attachment site, *attB* (in red), contain central regions of identity (I) (reviewed in refs 1, 2). Recombination between these central regions of *attP* and *attB* results in the formation of the CTX prophage with *attL* on its left border and *attR* on its right border. **b**, DNA sequences within *attP*, *attB*, *attL* and *attR*. The central sequences of identity, where recombination occurs, are shown in bold. Short segments of sequence surrounding the central identity region (phage sequences in blue and bacterial sequences in red) are shown as lower-case letters.

and not the CTX-Kn prophage (Fig. 2). Both mutant strains regained the ability to integrate phage DNA when wild-type Xer protein was produced from a complementing plasmid (Fig. 2).

As in *E. coli* *xer* mutants, cultures of *V. cholerae* strains harbouring mutations in either *xerC* or *xerD* contained a subpopulation of cells with a filamentous morphology (Fig. 3a–d). These filaments probably reflect aberrant cell division resulting from elimination of the *V. cholerae* Xer-mediated chromosome dimer resolution system^{6,8}. Filamentation in the *V. cholerae* *xer* mutants was eliminated by

complementation of the *xer* mutations *in trans* (Fig. 3e, f). To exclude the possibility that filamentation in *V. cholerae* *xer* mutants could affect CTX ϕ integration, we disrupted *recA* in the *V. cholerae* *xerC* and *xerD* mutant strains to suppress cell filamentation⁶. pCTX-Kn failed to integrate in these strain backgrounds (data not shown). This finding argues against the possibility that secondary consequences of the *xer* mutations, such as filamentation or SOS induction⁶, caused the failure of pCTX-Kn to form prophages in the *V. cholerae* *xerC* and *xerD* mutants.

In the 28-bp *E. coli* *dif* site, XerC and XerD bind to 11-bp half-sites separated by a 6-bp spacer of alternating pyrimidine and purine nucleotides (Fig. 4a)⁶. Inspection of the *V. cholerae* chrI CTX ϕ *attB* region revealed two overlapping sets of potential XerCD-binding sites, designated core-1 and core-2 in Fig. 4. Mutations in core-1 and core-2 were introduced into chrI of strain 2740-80 to investigate the contribution of these sequences to CTX ϕ integration and chromosome dimer resolution. Deletion of the entire 28-bp core-1 sequence in chrI, yielding 2740-80 Δ att1 (Fig. 4b, Δ att1), revealed that this region was required for both a functional CTX ϕ *attB* site (Fig. 2c) and a chrI *dif* site (Fig. 3g). Cell filamentation in 2740-80 Δ att1 was suppressed by a *recA* mutation (data not shown), a result consistent with the idea that the chrI core-1 sequence includes at least part of the chrI *dif* site⁸. CTX ϕ was able to integrate into 2740-80 Δ att2 (data not shown), a strain with a 9-bp deletion in the left end of core-1, including most of its putative XerC-binding site (Fig. 4b). XerCD recombination activity at *dif* was evidently intact in 2740-80 Δ att2 because cultures of this strain did not contain filamentous cells (Fig. 3h). CTX ϕ did not integrate into 2740-80 Δ att3 (data not shown), a strain that contains a 9-bp deletion in core-2 (Fig. 4b), including most of its putative XerD-binding site. The *dif* site was not functional in 2740-80 Δ att3 because cultures of this strain contained filamentous cells (Fig. 3i). Although the failure of CTX ϕ to integrate into 2740-80 Δ att3 indicates that part of core-2 is required for CTX ϕ integration, Xer-mediated recombination of the CTX ϕ *attP* with core-2 would not generate the *attL* and *attR* sequences described in Fig. 1. However, these sequences could be generated by recombination with core-1. Assuming that our delineation of the XerC- and XerD-binding sites is correct, these observations indicate that CTX ϕ *attP* recombines with core-1 and that core-2 is the chrI *dif* site.

We investigated whether the smaller *V. cholerae* chromosome (chrII) also contains a *dif* site. We have previously reported that chrII contains a 254-bp sequence that is almost identical to the chrI *attB* region⁹. We suspected that this region might contain chrII *dif*, as it lies at the approximate midpoint of chrII (ref. 8). Furthermore, in classical strains of *V. cholerae*, a CTX prophage is present upstream of this region⁹. Additional analysis of this sequence revealed that it contains two sets of putative XerCD-binding sites (Fig. 4a). Cultures of 2740-80 Δ att4, which contains a 37-bp deletion encompassing these chrII core-1 and core-2 sequences (Fig. 4b, Δ att4), had a subpopulation of filamentous cells (Fig. 3j), although filamentation was less pronounced than in the chrI *dif* deletion strains. Cell filamentation in 2740-80 Δ att4 was suppressed by a mutation in *recA* (data not shown). We conclude that *V. cholerae* chrII *dif* lies at the site of the 2740-80 Δ att4 deletion and that XerCD-mediated resolution of chromosome dimers at *dif* sequences is important for the segregation of both *V. cholerae* chromosomes. Because the chrII *dif* site does not serve as a CTX ϕ integration site in El Tor strains of *V. cholerae*, it is clear that phage integration and dimer resolution depend upon distinct sequences.

Comparison of the DNA sequences of the chrII *dif*, which is not a CTX ϕ integration site, with the chrI *dif/attB* site yielded clues to the sequence requirements for *attB* function. The core-2 sequences were highly conserved (Fig. 4), differing primarily in the spacer region; thus, it seems likely that on chrII, as on chrI, the core-2 region constitutes *dif*. In contrast, the two core-1 sequences were dissimilar (Fig. 4). Because the core-1 putative XerC-binding site on chrI was

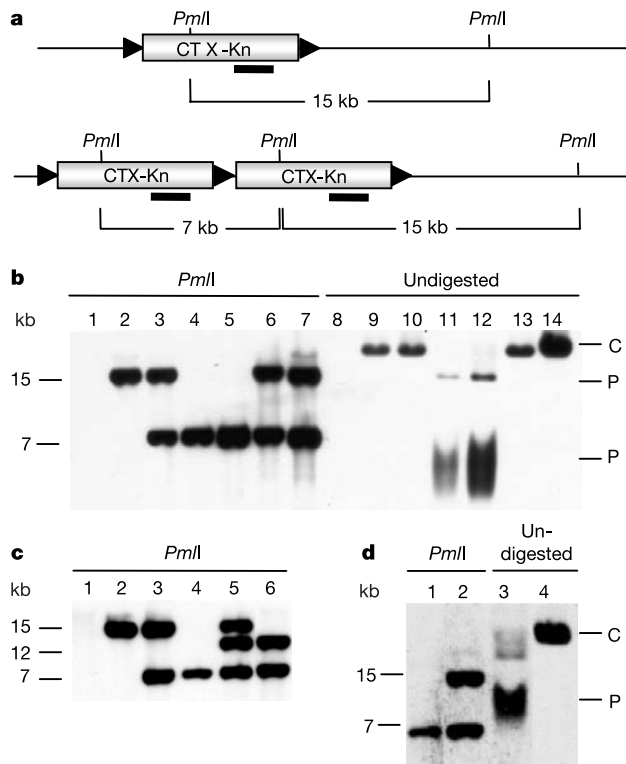


Figure 2 Analysis of pCTX-Kn integration in 2740-80 derivatives. pCTX-Kn was introduced into several 2740-80 derivatives and total DNA was prepared from Kn^R transformants. DNA samples were either digested with *Pml*, which cuts at a single site in pCTX-Kn, or left undigested. **a**, Schematic depiction of the predicted fragments detected in Southern analysis of *Pml*-digested 2740-80 DNA containing either a single CTX prophage (top) or two tandem CTX prophages (bottom). The site of binding of the *oriU* probe is shown below the prophages as a thick line. Undigested pCTX-Kn yields a 7-kb band after digestion with *Pml*. **b**, Southern blot analysis of pCTX-Kn integration into 2740-80 *xer* mutants. Total DNA was prepared from the following strains: 2740-80 (lanes 1 and 8); 2740-80(CTX-Kn) single prophage (lanes 2 and 9); 2740-80(CTX-Kn) two tandem prophages (lanes 3 and 10); 2740-80 *xerC*:Ap pCTX-Kn (lanes 4 and 11); 2740-80 *xerD*:Ap pCTX-Kn (lanes 5 and 12); 2740-80 *xerC*:Ap pXerC (CTX-Kn) two tandem prophages (lanes 6 and 13); and 2740-80 *xerD*:Ap pXerD (CTX-Kn) two tandem prophages (lanes 7 and 14). Only pCTX-Kn was detected in uncomplemented *xer* mutant strains (lanes 4, 5, 11 and 12), whereas the CTX-Kn prophage was observed in wild-type 2740-80 and in the complemented *xer* mutant strains (lanes 2, 3, 6, 7, 9, 10, 13 and 14). **c**, Southern blot analysis of pCTX-Kn integration into 2740-80 *dif/attB* mutants. Total DNA was prepared from the following strains: 2740-80 (lane 1); 2740-80(CTX-Kn) single prophage (lane 2); 2740-80(CTX-Kn) two tandem prophages (lane 3); 2740-80 Δ att1 pCTX-Kn (lane 4); 2740-80 *lacZ*::*attB* (CTX-Kn) (lane 5) and 2740-80 Δ att1 *lacZ*::*attB* (CTX-Kn) (lane 6). All strains contained the CTX-Kn prophage except 2740-80 Δ att1, which contained pCTX-Kn. **d**, Southern blot analysis of pCTX-Kn integration into an altered chrII *dif/attB* region in Bah2. Total DNA was prepared from strains Bah2 pCTX-Kn (lanes 1 and 3) and Bah2att7 (CTX-Kn) chrII prophage (lanes 2 and 4). In **b–d**, approximate sizes (kb) of the hybridizing fragments are shown to the left of the blots. Bands from uncut samples are identified at the right-hand side of the blots as either chromosomal DNA (C) or plasmid DNA (P).

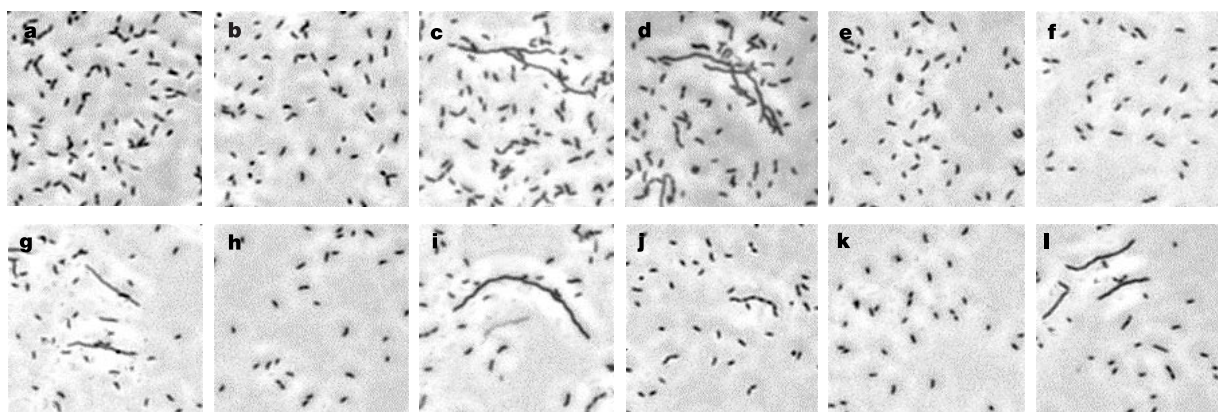


Figure 3 Filamentation in *V. cholerae* *xer* and *dif* mutants. Phase-contrast microscopy of cells (400 × magnification) from mid-logarithmic phase cultures of 2740-80 (a), 2740-80 (CTX-Kn) (b), 2740-80 *xerC*::Ap (c), 2740-80 *xerD*::Ap (d), 2740-80 *xerC*::Ap pXerC

(100 μM isopropyl-β-D-thiogalactoside (IPTG)) (e), 2740-80 *xerD*::Ap pXerD (100 μM IPTG) (f), 2740-80Δ*att1* (g), 2740-80Δ*att2* (h), 2740-80Δ*att3* (i), 2740-80Δ*att4* (j), 2740-80 *lacZ*::*attB* (k) and 2740-80Δ*att1* *lacZ*::*attB* (l).

not essential for CTXφ integration, we investigated whether a functional CTXφ *attB* could be generated on chrII solely by converting the putative chrII core-1 XerD-binding site to the sequence found in chrI core-1 (see bases with asterisks in Fig. 4b). These changes were introduced into Bah-2 (ref. 4), an El Tor strain that harbours a deletion of the chrI CTXφ integration site, resulting in strain Bah2att7. Ordinarily, only the plasmid form of CTXφ can be isolated from Bah-2 (ref. 3) (Fig. 2d). In contrast, CTXφ integrated into Bah2att7 (Fig. 2d), and the sequence of the prophage junctions in this strain revealed that integration occurred at the altered core-1 site in chrII (data not shown). This result supports the

hypothesis that core-1 (and in particular the putative XerD-binding site) is essential for CTXφ DNA integration.

In *E. coli*, resolution of chromosome dimers by XerCD occurs only if the *dif* site is within the chromosome's replication terminus¹⁰. Because CTXφ integration requires the XerCD recombinases and occurs at a site that overlaps *dif*, we tested whether the integration of CTXφ DNA was also restricted to the replication terminus of the *V. cholerae* chrI. For these experiments, a 350-bp fragment containing the chrI CTXφ *dif/attB* site was inserted into the *lacZ* locus, at about 300° on chrI (ref. 7) in *V. cholerae*. Kn^R transformants derived from introducing pCTX-Kn into 2740-80Δ*att1* *lacZ*::*attB* contained the CTX-Kn prophage in *lacZ* (Fig. 2c). When pCTX-Kn was introduced into 2740-80 *lacZ*::*attB*, CTX-Kn prophages were detected at both the *lacZ*::*attB* and native *attB* sites. Thus, integration can occur when *attB* is located outside the chrI replication terminus. However, cultures of 2740-80Δ*att1* *lacZ*::*attB* still contained a subpopulation of filamentous cells (Fig. 3k, l), indicating that, as in *E. coli*, proper location of the *V. cholerae* *dif* site is important for the resolution of chromosome dimers. Thus, the requirements for XerCD-mediated resolution at *dif* differ from those for CTXφ integration. Another notable difference between the two XerCD-mediated processes is that the carboxy terminus of FtsK is an essential accessory factor for chromosome dimer resolution¹¹, whereas it is dispensable for CTXφ integration (data not shown).

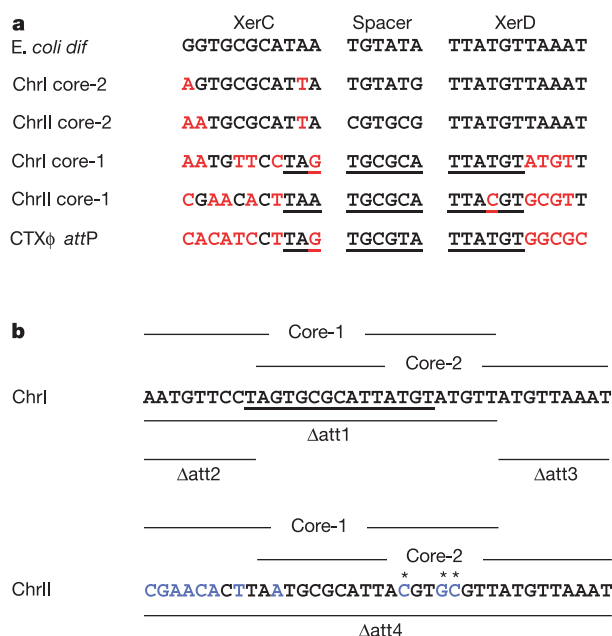


Figure 4 Sequences of the chrI and chrII *dif/attB* regions. **a**, Alignment of the *E. coli* chromosomal *dif* sequence with the sequences of the putative XerC- and XerD-binding sites (core regions) in the *V. cholerae* chrI and chrII CTXφ *dif/attB* and with the CTXφ *attP*. Nucleotides that differ from the *E. coli dif* sequence are highlighted in red. **b**, Putative overlapping sets of XerC- and XerD-binding sites within the chrI and chrII *dif/attB* regions (labelled core-1 and core-2). The boundaries of the 2740-80 chromosomal deletions are shown below the sequences. Nucleotides that differ between the chrI and chrII sequences are shown in blue. Asterisks indicate the three nucleotides altered in Bah2att7. In **a** and **b**, the sequences that recombine during the site-specific integration of CTXφ into chrI (central identity region in Fig. 1) are underlined.

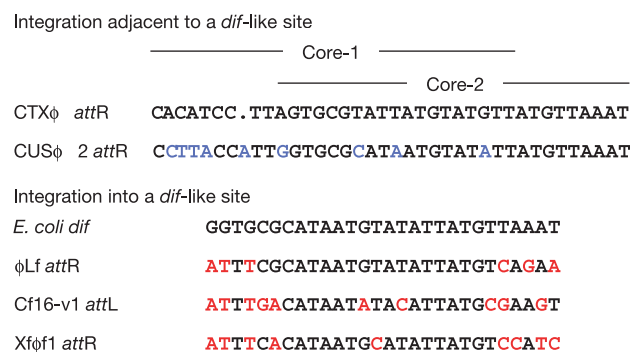


Figure 5 Alignment of *dif*-like chromosome junction sequences of various filamentous prophages. The sequences of *V. cholerae* CTXφ attR (Fig. 1), *Yersinia pestis* CUSφ-2 attR¹⁴, *Xanthomonas campestris* φLf attR¹⁵, *X. campestris* Cf16-v1 attL¹⁶ and *Xylella fastidiosa* Xfφf1 attR¹⁷ are shown. Core-1 and core-2 are as described in the legend to Fig. 4. Differences between CTXφ attR and CUSφ-2 attR are shown in blue and the differences between the *E. coli dif* and φLf attR, Cf16-v1 attL and Xfφf1 attR are shown in red.

The production of virion DNA by phages that integrate site-specifically into the host chromosome is usually preceded by prophage excision, which often requires the phage integrase¹. To test whether the formation of CTX ϕ virion DNA is similarly dependent on the XerCD recombinases, the *xerC* and *xerD* loci were inactivated in *V. cholerae* strains harbouring functional CTX-Kn prophages. The XerCD recombinases were not required for formation of CTX-Kn ϕ -transducing particles. These results are consistent with our model for the formation of CTX ϕ DNA from CTX prophages^{12,13}, in which prophage excision is not required for CTX ϕ DNA production.

Apparently, other filamentous phages have also exploited the XerCD recombination system to integrate into chromosomal *dif*-like sequences. Several filamentous prophages, including CUS-1, CUS-2 (E. Vimr, personal communication)¹⁴, ϕ Lf (ref. 15), Cfl6-v1 (ref. 16) and Xf ϕ 1, have chromosome junction sequences that resemble *dif* sequences (Fig. 5), indicating that these phage genomes become integrated into their host's chromosome by XerCD-mediated recombination. Xf ϕ 1 is a previously unannotated filamentous prophage that we have identified in the chromosome sequence of the plant pathogen *Xylella fastidiosa* 9a5a (ref. 17). In some cases, for example, CUS-1 and CUS-2, it seems that these phages integrate into a sequence containing overlapping sets of XerCD-binding sites. Other phage genomes, such as those of Cfl6-v1, ϕ Lf and Xf ϕ 1, seem to integrate into a sequence that contains a single set of XerCD-binding sites. Finally, filamentous phage might not be the only type of mobile genetic element captured by the XerCD recombinases. The recently identified pathogenicity island in *Neisseria gonorrhoeae* is flanked by *dif*-like sites¹⁸ (J. Dillard, personal communication). It is likely that these highly diverse horizontally transmitted elements have taken advantage of the XerCD recombination system because of its widespread distribution in eubacterial genomes and the conserved nature of the XerCD proteins and the *dif* recombination sites. □

Methods

Strain and plasmid construction

Derivatives of the suicide vector pGP704 containing internal fragments of the 2740-80 *xerC* and *xerD* genes were used for insertional inactivation of these genes¹⁹. The *xerC* internal fragment was amplified by polymerase chain reaction (PCR) with the primer pair XerCF1 (5'-AAGAACGACGAACGACCC-3')/XerCB1 (5'-TTGTGCTGTCCAGCAACAC-3') and the *xerD* internal fragment was amplified with the primer pair XerDF1 (5'-TGGCAGAAATACCGTGGCGTC-3')/XerDB1 (5'-TCACCCATAGGCACCAACG-3'). Disruptions of the target loci were verified by Southern analysis.

Complementing plasmids pXerC and pXerD were constructed by cloning the full-length *xerC* and *xerD* sequences amplified by PCR from 2740-80 chromosomal DNA with the primer pairs XerCF2 (5'-CCTTGGCAGTCACACGATG-3')/XerCB2 (5'-GCCAATGGCGGTAAACAGC-3') and XerDF2 (5'-GAGCTGAGGATAGAGTGAG-3')/XerDB2 (5'-CGCTCATGAAATCACC-3') into the expression vector pGZ119HE (ref. 20).

Deletion of *recA* from 2740-80 *xerC::Ap*, 2740-80 *xerD::Ap*, 2740-80 Δ att1 and 2740-80 Δ att4 was accomplished with the allele exchange vector pKM392, a Cm^R derivative of pGP84 (ref. 4). Mutations in the *dif/attB* sequences in chrI and chrII of 2740-80 were made by using derivatives of the allele exchange vector pWM91 (ref. 19) that had either the chrI or the chrII *dif/attB* regions with the desired mutations. These mutated *dif/attB* regions were made by overlap-extension PCR mutagenesis²¹ with 2740-80 chromosomal DNA as template and the outside primer pair DifA (5'-TCCATTACTGGTTGTCGG-3')/DifD (5'-AGAACTCCGCCTTAGTGTG-3') for chrI mutations and the primer pair Att4A (5'-TGAACGCAAGTCGCCGTAGAG-3')/Att4D (5'-TCCTCGTGAACCCCAACTGC-3') for the chrII mutations. The inside primer pairs were as follows (the overlapping sequences within the primers are underlined): for Δ att1, DifB (5'-AGAGCGCGGGCGGCATTGTCTCTAGGTTTA-3')/DifC (5'-TGCCGCCCGCGCTCTATGTTAAATTAAGGC-3'), for Δ att2, DifB/Att2C (5'-TGCCGCCCGCGCTCTAGTGCCTATGTTATG-3'), for Δ att3, Att3B (5'-AGAGCGCGGGCGGCATAACATACATAATGCGCAG-3')/Att3C (5'-TGCCGCCCGCGCTCTAAGGCATAAAAGAGGTGCG-3'); for Δ att4, Att4B (5'-AGAGCGCGGGCGGCATAGCATTTCTAATCATC-3')/Att4C (5'-TGCCGCCCGCGCTCTAAGGCATAAAAGAGG-3'); and for Att7, Att7B (5'-CTAGGAACATTCTAATCATCACTCGTGACG-3')/Att7C (5'-TTAGAATGTTCCATGTCGCTATGTTATGTTATG-3'). All chromosomal mutations were verified by automated DNA sequencing performed at the Tufts Core Facility.

A DNA fragment containing the chrI *dif/attB* site was inserted into the 2740-80 *lacZ* locus by allele exchange by using a derivative of pJL1 (ref. 22) containing a 350-bp PCR

product including the chrI *dif/attB* region amplified from 2740-80 with the primer pair AttB1 (5'-CGCAGCAGACGAACCTCTATGTC-3')/AttB2 (5'-CACTTGTGTGCACACACAATTGACG-3').

CTX ϕ integration assay

pCTX-Kn was introduced into *V. cholerae* strains by electroporation; total DNA was then isolated with the Genome DNA kit (BIO 101, Carlsbad, California, USA) from Kn^R transformants. The samples were analysed by Southern blotting with an *orfU* DNA probe by using the ECL direct labelling system (Amersham-Pharmacia Biotech, Little Chalfont, UK). This probe was made by PCR amplification of pCTX-Kn template DNA and primers OrfU1F (5'-ATGCGCTATTTTCTACTGTTTGTG-3') and OrfU4R (5'-TCTAGAAAATCACCTAAACAAAATGAGCATG-3'). All assays were performed on at least three independently derived transformants. CTX ϕ titres were measured as described previously³.

Received 20 February; accepted 18 March 2002; doi:10.1038/nature00782.

- Nash, H. A. in *Escherichia coli* and *Salmonella* (ed. Neidhardt, F. C.) 2363–2376 (ASM Press, Washington DC, 1996).
- Campbell, A. M. Chromosomal insertion sites for phages and plasmids. *J. Bacteriol.* **174**, 7495–7499 (1992).
- Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914 (1996).
- Pearson, G. D., Woods, A., Chiang, S. L. & Mekalanos, J. J. CTX genetic element encodes a site-specific recombination system and an intestinal colonization factor. *Proc. Natl Acad. Sci. USA* **90**, 3750–3754 (1993).
- Davis, B. M., Kimsey, H. H., Chang, W. & Waldor, M. K. The *Vibrio cholerae* O139 Calcutta bacteriophage CTX ϕ is infectious and encodes a novel repressor. *J. Bacteriol.* **181**, 6779–6787 (1999).
- Blakely, G. et al. Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell* **75**, 351–361 (1993).
- Heidelberg, J. F. et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477–483 (2000).
- Kuempel, P. L., Henson, J. M., Dircks, L., Tecklenburg, M. & Lim, D. F. *dif*, a *recA*-independent recombination site in the terminus region of the chromosome of *Escherichia coli*. *New Biol.* **3**, 799–811 (1991).
- Davis, B. M., Moyer, K. E., Boyd, E. F. & Waldor, M. K. CTX prophages in classical biotype *Vibrio cholerae*: functional phage genes but dysfunctional phage genomes. *J. Bacteriol.* **182**, 6992–6998 (2000).
- Tecklenburg, M., Naumer, A., Nagappan, O. & Kuempel, P. The *dif* resolvase locus of the *Escherichia coli* chromosome can be replaced by a 33-bp sequence, but function depends on location. *Proc. Natl Acad. Sci. USA* **92**, 1352–1356 (1995).
- Steiner, W., Liu, G., Donachie, W. D. & Kuempel, P. The cytoplasmic domain of FtsK protein is required for resolution of chromosome dimers. *Mol. Microbiol.* **31**, 579–583 (1999).
- Moyer, K. E., Kimsey, H. H. & Waldor, M. K. Evidence for a rolling-circle mechanism of phage DNA synthesis from both replicative and integrated forms of CTX ϕ . *Mol. Microbiol.* **41**, 311–323 (2001).
- Davis, B. M. & Waldor, M. K. CTX ϕ contains a hybrid genome derived from tandemly integrated elements. *Proc. Natl Acad. Sci. USA* **97**, 8572–8577 (2000).
- Parkhill, J. et al. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527 (2001).
- Lin, N. T., Chang, R. Y., Lee, S. J. & Tseng, Y. H. Plasmids carrying cloned fragments of RF DNA from the filamentous phage ϕ Lf can be integrated into the host chromosome via site-specific integration and homologous recombination. *Mol. Genet. Genom.* **266**, 425–435 (2001).
- Dai, H., Chow, T. Y., Liao, H. J., Chen, Z. Y. & Chiang, K. S. Nucleotide sequences involved in the neolysogenic insertion of filamentous phage Cfl6-v1 into the *Xanthomonas campestris* pv. citri chromosome. *Virology* **167**, 613–620 (1988).
- Simpson, A. J. et al. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**, 151–157 (2000).
- Dillard, J. P. & Seifert, H. S. A variable genetic island specific for *Neisseria gonorrhoeae* is involved in providing DNA for natural transformation and is found more often in disseminated infection isolates. *Mol. Microbiol.* **41**, 263–277 (2001).
- Metcalfe, W. W. et al. Conditionally replicative and conjugative plasmids carrying *lacZ* alpha for cloning, mutagenesis, and allele replacement in bacteria. *Plasmid* **35**, 1–13 (1996).
- Lessl, M. et al. Dissection of IncP conjugative plasmid transfer: definition of the transfer region Tra2 by mobilization of the Tra1 region in *trans*. *J. Bacteriol.* **174**, 2493–2500 (1992).
- Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61–68 (1989).
- Butterton, J. R. et al. Heterologous antigen expression in *Vibrio cholerae* vector strains. *Infect. Immun.* **63**, 2689–2696 (1995).

Acknowledgements

We thank L. Arciszewska, A. Camilli, A. Campbell, B. Davis, A. Kane, D. Raychaudhuri, M. Russel and A. Sonenshein for helpful suggestions, and E. Vimr for communication of unpublished observations. We are grateful to A. Kane and the New England Medical Center GRASP Center for the preparation of plates and media. We acknowledge the support of NIH, Howard Hughes Medical Institute and the Pew Foundation.

Competing interests statement

The authors declare that they have no competing financial interests

Correspondence and requests for materials should be addressed to M.K.W. (e-mail: mwaldor@lifespan.org).

A conserved RNA-binding protein controls germline stem cells in *Caenorhabditis elegans*

Sarah L. Crittenden*, David S. Bernstein†, Jennifer L. Bachorik‡, Beth E. Thompson‡, Maria Gallegos‡§, Andrei G. Petcherski†§, Gary Moulder||, Robert Barstead||, Marvin Wickens† & Judith Kimble*†

* Howard Hughes Medical Institute; † Department of Biochemistry; and ‡ Program in Cellular and Molecular Biology, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA
§ Program in Molecular Biology, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104, USA

Germline stem cells are defined by their unique ability to generate more of themselves as well as differentiated gametes¹. The molecular mechanisms controlling the decision between self-renewal and differentiation are central unsolved problems in developmental biology with potentially broad medical implications. In *Caenorhabditis elegans*, germline stem cells are controlled by the somatic distal tip cell^{2,3}. FBF-1 and FBF-2, two nearly identical proteins, which together are called FBF ('*fem-3* mRNA binding factor'), were originally discovered as regulators of germline sex determination⁴. Here we report that FBF also controls germline stem cells: in an *fbf-1 fbf-2* double mutant, germline proliferation is initially normal, but stem cells are not maintained. We suggest that FBF controls germline stem cells, at least in part, by repressing *gld-1*, which itself promotes commitment to the meiotic cell cycle^{5,6}. FBF belongs to the PUF family ('Pumilio and FBF') of RNA-binding proteins⁷. Pumilio controls germline stem cells in *Drosophila* females^{8,9}, and, in lower eukaryotes, PUF proteins promote continued mitoses^{10,11}. We suggest that regulation by PUF proteins may be an ancient and widespread mechanism for control of stem cells.

In the *C. elegans* germ line, mitotic cells reside distally and differentiating gametes are proximal (Fig. 1a). The mitotic region includes germline stem cells. By analogy with other tissues, the mitotic region may also contain cells with a more limited capacity to proliferate, sometimes called transit amplifying cells¹. Germline mitotic divisions are controlled by the somatic distal tip cell (DTC), which uses Notch signalling to promote mitosis and prevent meiosis^{2,3}. Therefore, the DTC provides a niche for maintaining germline stem cells. Once germ cells reach the transition zone, they enter meiosis and then progress through the pachytene stage of meiotic prophase and gametogenesis as they move proximally^{3,12}. Downstream of Notch signalling, the *gld-1* and *gld-2* genes promote commitment to the meiotic cell cycle^{5,6,13}.

The *fbf-1* and *fbf-2* genes encode nearly identical (>91%) proteins that bind specifically to an RNA regulatory element in the *fem-3* 3' untranslated region (3' UTR), and thereby promote the hermaphroditic switch from spermatogenesis to oogenesis⁴. To investigate FBF function in more depth, we generated an *fbf-1* single mutant and then induced an *fbf-2* mutation on the *fbf-1* chromosome to generate an *fbf-1 fbf-2* double mutant (Fig. 1b). The *fbf-1* deletion removes the RNA-binding region and alters the reading frame (Fig. 1b, top); *fbf-2*(q704) is a nonsense mutation (Fig. 1b, bottom). Both mutations are likely to abolish FBF activity. Consistent with this idea, similar results have been obtained using another single mutant, *fbf-1(ok224)* and another double mutant, *fbf-1(q662) fbf-2(q655)*. Furthermore, RNA-mediated interference (RNAi) of either *fbf-1* or *fbf-2* targets both *fbf* messenger RNAs and

has the same effect as the double mutant (not shown).

The phenotypes of the *fbf-1* single mutant and the *fbf-1 fbf-2* double mutant suggest that FBF-1 and FBF-2 are largely redundant. Thus, *fbf-1* mutant hermaphrodites were fertile and virtually wild type, although they made more sperm than normal (Fig. 1c). By contrast, *fbf-1 fbf-2* double mutants were sterile with severe germline defects (Fig. 1c, e, g, h). The *fbf-1 fbf-2* double mutant had two prominent germline defects. First, no switch from spermatogenesis to oogenesis occurred, as predicted from RNAi experiments using single-stranded RNA⁴. Second, all cells in the mitotic region entered meiosis in late stage 4 larvae (L4). Early germline development appeared normal: newly hatched animals possessed two germline precursor cells in wild type³ and in the double mutant ($n = 19$). Early germline divisions generated an average of 31 descendants by L3 in wild type ($n = 10$, range 26–38), and 32 descendants by the

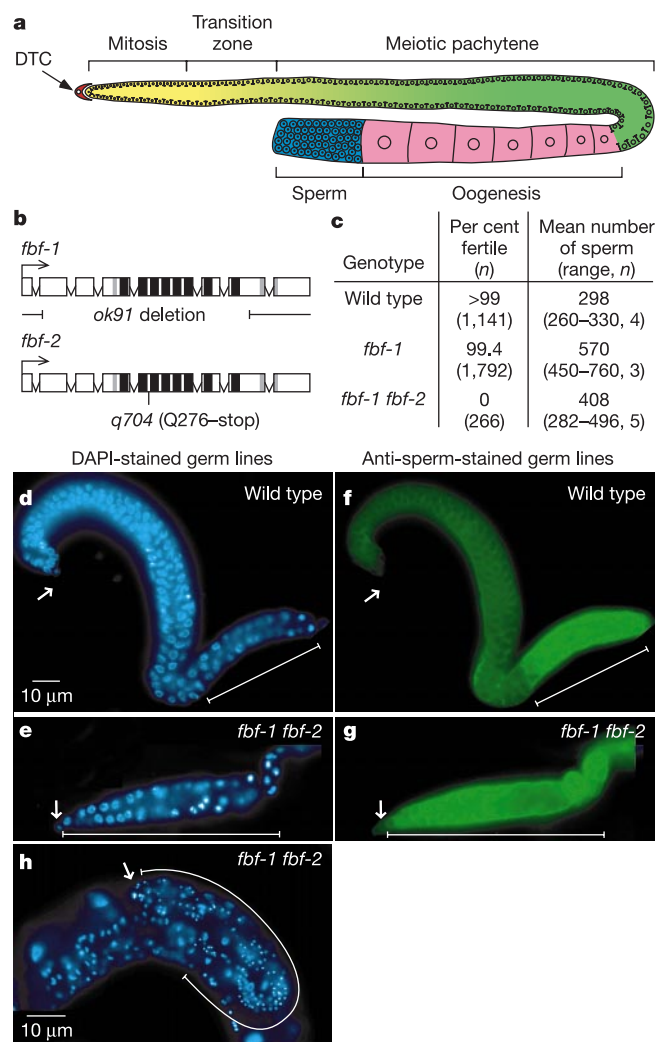


Figure 1 FBF is required for germline stem cells. **a**, Germline organization is the same in both sexes, except hermaphrodites make sperm and oocytes, whereas males make only sperm. **b**, Molecular lesions in *fbf* mutations. Black boxes, Puf repeats, which govern RNA binding; grey, Csp region⁴. **c**, Fertility in *fbf* mutants. n , number animals scored. **d–g**, L4 hermaphrodite germ lines stained with 4,6-diamidino-2-phenylindole (DAPI) and anti-sperm as described²³. **d** and **f** are same germ line, as are **e** and **g**, and all are at the same magnification. **h**, Adult stained with DAPI as described⁶. Arrow, distal end; bar, spermatogenic region.

§ Present addresses: Department of Anatomy, The University of California, San Francisco, California 94143, USA (M.G.); Wormbase, Caltech 156-29, Pasadena, California 91125, USA (A.G.P.)

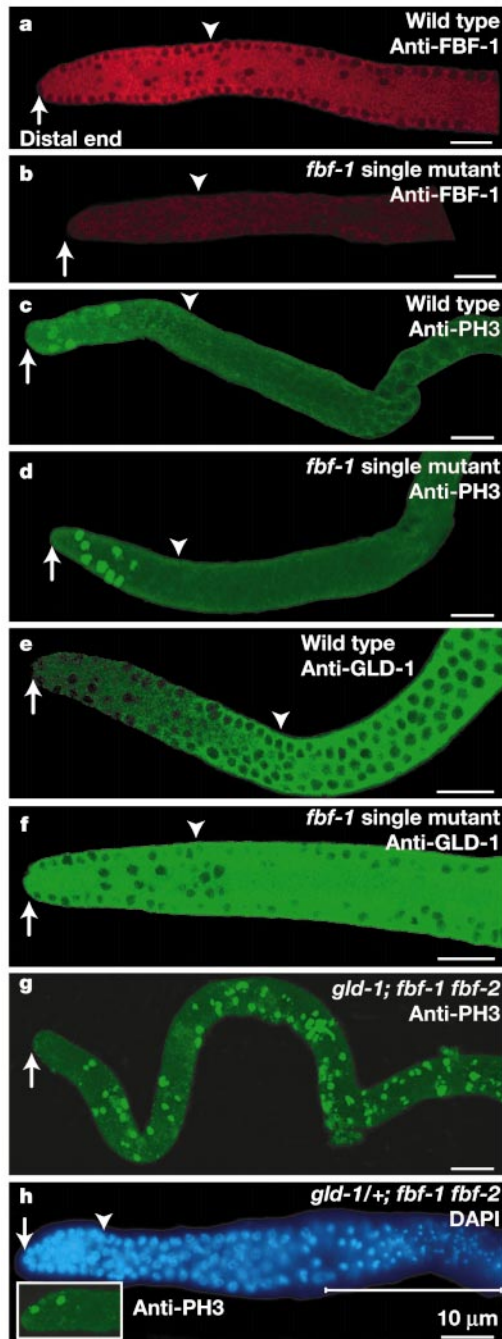


Figure 2 FBF-1 controls GLD-1 expression. Adult hermaphroditic gonads were stained²³ with anti-FBF-1 (**a, b**), anti-phosphohistone H3 (anti-PH3, Upstate Biotechnology), which stains dividing cells²⁴ (**c, d, g, h**), anti-GLD-1 (ref. 14) (**e, f**) or DAPI (**h**). Arrow, distal end; arrowhead, approximate end of mitotic region. **a**, FBF-1 in wild-type germ line. **b**, FBF-1 in *fbf-1* mutant; no staining seen. **c**, Mitotic region in wild type. **d**, Mitotic region in *fbf-1* mutant. Nuclei positive for anti-PH3 (green) can extend 16–20 germ-cell diameters from the distal end in wild-type ($n = 21$) and 12–14 cell diameters in the mutant ($n = 21$). **e**, Wild type. GLD-1 is low in distal-most cells ($n = 17$; also see ref. 14). **f**, *fbf-1* mutant. High GLD-1 extends to the distal end ($n = 16$ germ lines). **g**, The *gld-1(q485); fbf-1(ok91) fbf-2(q704)* germ line has many nuclei positive for anti-PH3 ($n = 16$ of 16). **h**, The *gld-1(q485) + ; fbf-1(ok91) fbf-2(q704)* germ line has a restored distal mitotic region with anti-PH3-positive nuclei (inset) and makes mature sperm. White bar, spermatogenic region. All scale bars, 10 μ m.

same stage in the double mutant ($n = 10$, range 26–39). However, by L4, wild-type germ lines continued to proliferate (Fig. 1d), but all *fbf-1 fbf-2* germ cells had entered meiotic pachytene (Fig. 1e). Furthermore, whereas the wild-type germ line had a restricted region of spermatogenesis (Fig. 1f, green), all *fbf-1 fbf-2* germ cells entered spermatogenesis (Fig. 1g, green). No morphologically aberrant or dying germ cells were seen at any stage ($n > 10$, each stage). By adulthood, the *fbf-1 fbf-2* germ line consisted entirely of mature sperm (Fig. 1h). On average, each of the two gonadal arms made 50 germ cells (range 35–60; $n = 10$ arms), or a total of 100 germ cells per animal, making about 400 mature sperm (Fig. 1c). By contrast, wild-type adult hermaphrodites have about 2,000 total germ cells and about 300 mature sperm^{2,3}. Therefore, the double mutant fails to switch into oogenesis and fails to maintain germline stem cells. A similar defect was observed in *fbf-1 fbf-2* mutant males (not shown). We conclude that FBF controls the maintenance of germline stem cells, but is not required for early larval germline proliferation.

Previous work showed FBF-1 in the L3 germline cytoplasm⁴. We now report that FBF-1 is present uniformly in germline cytoplasm during early larval stages, and becomes enriched in the mitotic region of L4 larval (not shown) and adult germ lines (Fig. 2a). There is no detectable staining in an *fbf-1* single mutant (Fig. 2b). Within the mitotic region of wild-type germ lines, the distal-most germ cells possess faint but easily detectable FBF-1 staining, which becomes intense in more proximal mitotic cells (Fig. 2a). As germ cells leave the mitotic region and enter early meiotic prophase, FBF-1 is reduced to a low level. This pattern is similar in males (not shown). Thus FBF-1 is found in mitotic germline cells in L4 larvae and adults, when FBF is required to maintain germline stem cells.

One mechanism by which FBF could maintain germline stem cells is to inhibit activity of a gene controlling commitment to meiosis. One such gene is *gld-1* (refs 5, 6). Using an affinity-purified antibody to GLD-1 (ref. 14), we first examined GLD-1 in an *fbf-1* single mutant. We chose the single mutant because the *fbf-1 fbf-2* double mutant dramatically changes germline cell fates, making the analysis of any changes in GLD-1 expression difficult to interpret. By contrast, *fbf-1* single mutants are nearly wild type with respect to germline fates and pattern, including the mitotic region (Fig. 2c, d). In wild-type germ lines, GLD-1 was either expressed at a low level or was undetectable in the distal-most germ cells (Fig. 2e)¹⁴. However, in *fbf-1* single-mutant germ lines, GLD-1 was detectable in all distal germ cells (16 of 18 germ lines), and abundant in most of them (12 of 18 germ lines) (Fig. 2f). This distal expansion of GLD-1 in the *fbf-1* mutant is consistent with the idea that FBF negatively regulates GLD-1 expression. The fact that the distal expansion of GLD-1 is not sufficient to drive germline stem cells into meiosis may reflect the fact that *fbf-2* remains or that other regulators are required in addition to GLD-1 (see below).

To investigate whether FBF acts genetically upstream of *gld-1*, we examined a *gld-1; fbf-1 fbf-2* triple mutant. In wild-type germ lines, mitotic cells are restricted to the distal region (Fig. 1a, 2c)², whereas in *gld-1* null mutants, proximal germ cells also divide mitotically⁵. The *gld-1; fbf-1 fbf-2* triple mutants displayed mitotically dividing cells throughout the germ line (Fig. 2g); in particular, distal germ cells were mitotic at all stages examined (early L4, late L4 and adult). Therefore, in the absence of *gld-1*, FBF is no longer necessary to promote germline proliferation. This epistasis is consistent with FBF acting as a negative regulator of *gld-1*. We also examined *fbf-1 fbf-2* double mutants that were heterozygous at the *gld-1* locus (*gld-1/+*), and were surprised to find that the germ line regained a distal mitotic region (Fig. 2h). Indeed, these germ lines were considerably larger than in the *fbf-1 fbf-2* double mutant (270 germ cells per arm compared with 50 germ cells per arm) and made about 900 mature sperm per animal ($n = 4$), more than twice as many sperm as an *fbf-1 fbf-2* double mutant. Therefore, the distal

mitotic cells were both proliferating and generating gametes, two defining features of stem cells. Thus reduction of *gld-1* by one copy seems to restore germline stem cells to the *fbf-1 fbf-2* double mutant, which strongly supports the idea that *gld-1* repression is critical for the maintenance of germline stem cells.

We hypothesized that FBF-1 and FBF-2 repress *gld-1* expression by binding 3'UTR elements, and therefore examined the *gld-1* 3'UTR sequence for possible FBF-binding sites. FBF binding

requires a UGUR tetranucleotide and adjacent sequences (D.B. and M.W., manuscript in preparation). Two potential FBF-binding elements (FBEs) were identified: FBE-a and FBE-b (Fig. 3a). To test FBF binding, we first used the yeast three-hybrid system¹⁵ (Fig. 3b). FBF-1 and FBF-2 bound both sites, activating *lacZ* (Fig. 3c) and *HIS3* reporter genes (Fig. 3d, and not shown). Indeed, FBF-1 binding to FBE-a is one of the strongest interactions yet detected in the system (Fig. 3d)¹⁶. FBF–FBE interactions were specific by three criteria. Mutations in the UGU of FBE-a* and FBE-b* (Fig. 3a) disrupted binding (Fig. 3c–e); FBF did not bind other RNAs, including several that bind other PUF proteins (Fig. 3c, d, and not shown); and a distinct *C. elegans* protein, PUF-5, failed to bind either site (Fig. 3d). The FBF–FBE interactions were direct, as judged by gel retardation assays (Fig. 3e). Purified, recombinant FBF-1 labelled with glutathione S-transferase (GST) bound ³²P-radiolabelled FBE-a and FBE-b RNAs carrying wild-type, but not mutant, FBEs (Fig. 3e). We conclude that FBF binds specifically to two sites in the *gld-1* 3'UTR.

We have provided cytological, genetic and molecular evidence that FBF represses *gld-1* mRNA activity. A fourth line of evidence relies on *gld-1(oz10)*, which deletes 500 nucleotides from the *gld-1* 3'UTR, including FBE-a and FBE-b (Fig. 3a)¹⁷. Interpretation is complicated, because the mutant also carries a missense mutation¹⁷. Nonetheless, GLD-1 expression expands to the distal end in *gld-1(oz10)* mutants¹⁴, supporting the hypothesis that the FBE sites mediate *gld-1* repression. The *gld-1* mRNA encodes a translational repressor of the STAR/Quaking/GSG family^{17,18}. It is possible that PUF proteins repress such mRNAs more generally.

FBF repression of *gld-1* maintains mitotically dividing germline cells, including stem cells. In *fbf-1 fbf-2* double mutants, all germline cells enter meiosis, whereas in *gld-1; fbf-1 fbf-2* triple mutants, mitoses are found throughout the germ line. Remarkably, removal of even one wild-type copy of the *gld-1* gene seems to restore stem cells to the *fbf-1 fbf-2* double mutant. Therefore, low GLD-1 activity is essential for germline stem cells. Yet we have not been able to show the converse: high GLD-1 may not be sufficient to drive germ cells into meiosis. One explanation is that the increased GLD-1 observed in *fbf-1* and *gld-1(oz10)* mutants does not represent fully unregulated wild-type protein. In *fbf-1* mutants, FBF-2 remains, and in

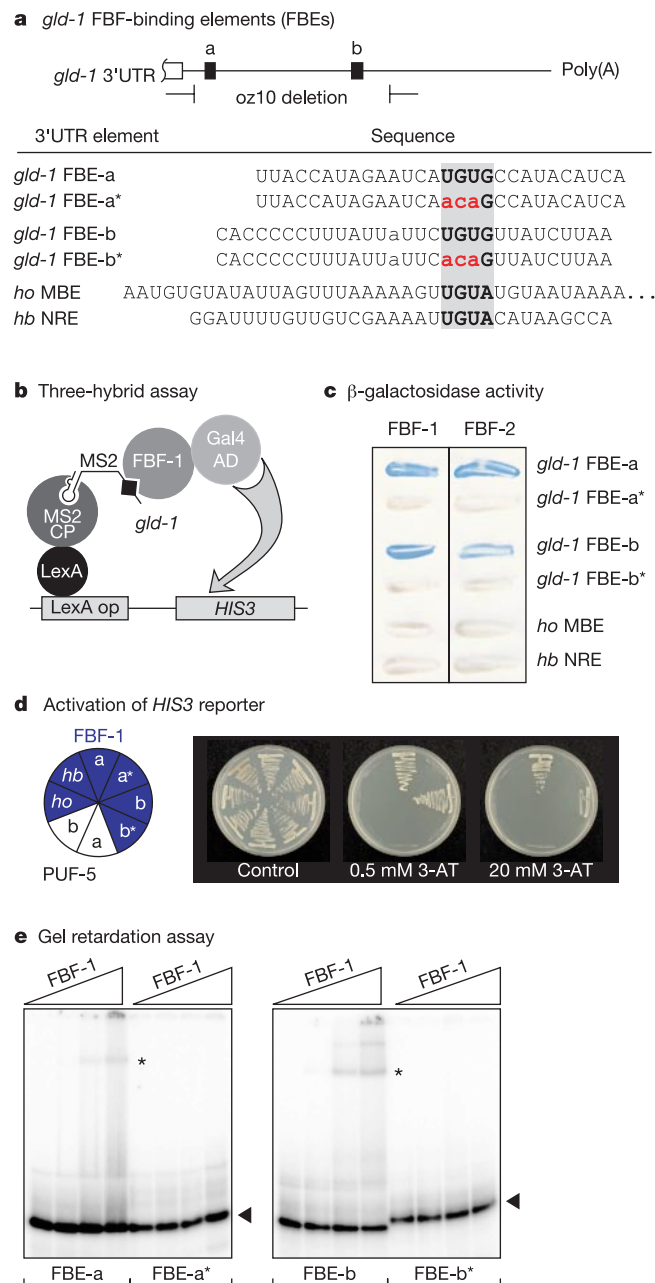


Figure 3 FBF binds specifically to the *gld-1* 3'UTR. **a**, Predicted FBF-binding elements (FBEs) in *gld-1* 3'UTR (FBE-a and FBE-b, 896 and 498 nt upstream of polyadenylation site, respectively). The UGUR motif is common among PUF-binding sites (grey box). Red letters are point mutations. Controls, *hunchback* (*hb*) Nanos response element (NRE)²⁵ and *ho* MPT-5/PUF-5 element (MBE)²⁶. **b**, Three-hybrid assay^{15,16}. **c**, β -galactosidase activity. **d**, *HIS3* activation. Left, letters indicate hybrid RNAs ('a' is *gld-1* FBE-a, and so on). Cells expressing FBF-2 (blue sectors) or *C. elegans* PUF-5 (white sectors). Right, growth was monitored on media lacking histidine and containing 3-AT, a competitive inhibitor of *HIS3*. **e**, Gel retardation assays. ³²P-RNAs were incubated with 0, 50, 250 or 500 pmol GST–FBF-1. Asterisk, specific complex; arrowhead, free RNA.

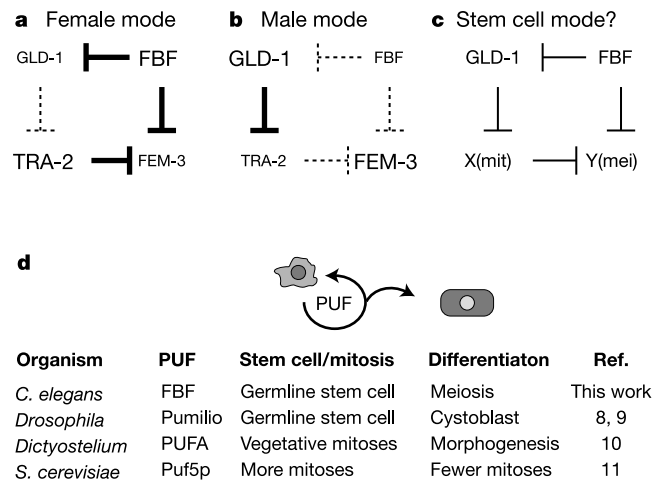


Figure 4 PUF protein regulation of cell fates. **a**, **b**, Regulation of sex determination. Large letters refer to active regulators; smaller letters refer to repressed regulators. Thick bars represent active repression; dashed bars represent lack of repression. **c**, Regulation of mitosis and meiosis. Letters and bars are given equal strength since this circuit is speculative. See text for further explanation. **d**, Role of PUF proteins in the nematode worm *C. elegans*, the fruitfly *Drosophila melanogaster*, the ameboid protozoan *Dictyostelium discoideum*, and the budding yeast *Saccharomyces cerevisiae*.

gld-1(oz10), GLD-1 protein carries a missense defect¹⁷. In addition, other regulators (such as GLD-2; ref. 6) may be required to drive germ cells into meiosis.

FBF provides a direct link between two major germline decisions: maintenance of stem cells (this work) and germline sex determination⁴. Intriguingly, GLD-1 provides a second and opposite link between these same decisions^{5,6,13,18}. Thus, FBF promotes mitosis and oogenesis, whereas GLD-1 promotes meiosis and spermatogenesis. We suggest the existence of a multiply reinforced translational switch that controls germline fate decisions. The circuitry is best understood for sex determination, where a series of negative regulators have been identified³ (Fig. 4a, b). A similar circuit may also control the mitosis/meiosis decision (Fig. 4c). Although some GLD-1 target mRNAs have been identified^{19,20}, their role in stem cell control remains unknown. Identification of FBF and GLD-1 target mRNAs (X(mit) and Y(me)) in Fig. 4c) is needed to elucidate how the circuit controls the decision between mitosis and meiosis.

Our results, together with those of others, suggest a conserved role for PUF proteins (Fig. 4d). FBF control of germline stem cells in *C. elegans* is reminiscent of the Pumilio control of germline stem cells in *Drosophila* ovaries^{8,9}. Remarkably, PUF proteins in yeast and the protozoan *Dictyostelium* also promote mitosis at the expense of an alternate state (Fig. 4d). In *Dictyostelium*, PufA mutants cease vegetative mitotic divisions, aggregate and undergo morphogenesis prematurely¹⁰. In yeast, PUF5/MPT5/UTH4 mutants divide fewer times and age early, whereas cells over-expressing this protein divide more times and age late¹¹. The common theme is that PUF proteins promote cell division at the expense of differentiation. This conservation suggests that additional components of the circuitry may also be conserved, an idea that awaits identification of PUF targets critical for stem cell maintenance in multiple organisms. □

Methods

Mutants

The *fbf-1(ok91)* deletion mutant was isolated in a screen based on polymerase chain reaction (PCR) as described²¹. Primers were: 201A, GTCAACGAGAGGAAATCTTCG; 202S, CCAGTGGCCATAATCGTGTG; 203A, GCGTAATGAATATTTTAGTGTG; 204S, CCTGAATAATGATTGTATTCTC. To obtain the *fbf-1 fbf-2* double mutant, *fbf-1(ok91)* males were mutagenized with ethyl methane sulphonate and crossed into *mIn1[mIs14 dpy-10(e128)]* homozygotes, which carry a green fluorescent protein (GFP)-marked balancer for chromosome II obtained from M. Edgley. Non-green progeny from *fbf-1*/mIn1* heterozygotes were screened for sterility. Potential doubles were tested by complementation to a previously isolated double mutant and then sequenced to determine the *fbf-2* lesion. Each mutant was outcrossed against wild type at least six times. *fbf-1(ok91)* was maintained as a homozygous stock. *fbf-1(ok91) fbf-2(q704)* homozygotes were offspring of *fbf-1 fbf-2/mIn1[mIs14 dpy-10(e128)]* hermaphrodites. *gld-1; fbf-1 fbf-2* and *gld-1/+; fbf-1 fbf-2* mutants were progeny from *gld-1(q485)/ccls4251 unc-15; fbf-1(ok91) fbf-2(q704)/mIn1[mIs14 dpy-10(e128)]* hermaphrodites.

Staining

Anti-FBF-1 antibodies⁴ are specific for FBF-1: they do not recognize FBF-2 on western blots of protein translated *in vitro* (not shown) or in stained gonads from *fbf-1* deletion mutants (Fig. 2b). Affinity-purified anti-GLD-1 (ref. 14) was a gift from T. Schedl. The sperm-specific monoclonal antibody, SP56 (ref. 22), was a gift from S. Ward. Images were obtained on a Bio-Rad MRC1024 confocal microscope or a Hamamatsu Orca camera with Openlabs software and processed using Adobe Photoshop. Images of FBF-1 in wild-type and *fbf-1* mutant backgrounds were processed identically, as were those of GLD-1 in the same backgrounds. To visualize dividing nuclei, we projected a z-series so that all nuclei positive for anti-phosphohistone H3 antibodies (anti-PH3) from a single germ line were visible.

Three-hybrid and gel retardation assays

Three-hybrid assays were performed as described¹⁶. The entire RNA sequences indicated in Fig. 3a were cloned into the *XmaI* and *SphI* sites of the vector pIII/MS2-2, using annealed synthetic oligonucleotides. For the three-hybrid assay, a mutation (U to A, indicated by lower case in Fig. 3a) was introduced into *gld-1*-FBE-b to disrupt an oligo(U) tract that can cause RNA polymerase III termination in the three-hybrid assay. The RNA-binding domains of FBF-1 (amino acids 121–614), FBF-2 (amino acids 2–632) and PUF-5 (amino acids 1–553, Genbank accession number NM_063413) were cloned into pACT2. Plasmids encoding the hybrid RNA and chimaeric protein were co-transformed into strain YBZ-1 (ref. 16). Levels of 3-aminotriazole (3-AT) resistance were determined by assaying

multiple transformants at ten different concentrations of 3-AT, up to 100 mM.

β -galactosidase activities were determined using colony filter assays¹⁶. *In vitro* binding reactions (Fig. 3e) contained 50 pmol of ³²P-RNA and 0–500 pmol of protein, in 100 mM NaCl, 5 mM dithiothreitol, 5 mM MgCl₂, 2 mM HEPES buffer at pH 7.4, and 200 ng yeast RNA. RNA and protein were incubated at room temperature for 15 min, heparin was added to 10 mg ml⁻¹, and the mixture was run at 4 °C on a 6% native polyacrylamide gel that contained 0.5 × TBE with 5 mM MgCl₂.

Received 8 January; accepted 6 May 2002; doi:10.1038/nature754.

Published online 22 May 2002.

- Watt, F. M. & Hogan, B. L. Out of Eden: stem cells and their niches. *Science* **287**, 1427–1430 (2000).
- Kimble, J. E. & White, J. G. On the control of germ cell development in *Caenorhabditis elegans*. *Dev. Biol.* **81**, 208–219 (1981).
- Schedl, T. in *C. elegans II* (ed. Priess, J. R.) 241–269 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1997).
- Zhang, B. *et al.* A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* **390**, 477–484 (1997).
- Francis, R., Barton, M. K., Kimble, J. & Schedl, T. *gld-1*, a tumor suppressor gene required for oocyte development in *Caenorhabditis elegans*. *Genetics* **139**, 579–606 (1995).
- Kadyk, L. C. & Kimble, J. Genetic regulation of entry into meiosis in *Caenorhabditis elegans*. *Development* **125**, 1803–1813 (1998).
- Wickens, M., Bernstein, D. S., Kimble, J. & Parker, R. A PUF family portrait: 3' UTR regulation as a way of life. *Trends Genet.* **18**, 150–157 (2002).
- Lin, H. & Spradling, A. C. A novel group of *pumilio* mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* **124**, 2463–2476 (1997).
- Forbes, A. & Lehmann, R. Nanos and Pumilio have critical roles in the development and function of *Drosophila* germline stem cells. *Development* **125**, 679–690 (1998).
- Souza, G. M., da Silva, A. M. & Kuspa, A. Starvation promotes *Dictyostelium* development by relieving PufA inhibition of PKA translation through the YakA kinase pathway. *Development* **126**, 3263–3274 (1999).
- Kennedy, B. K. *et al.* Redistribution of silencing proteins from telomeres to the nucleolus is associated with extension of life span in *S. cerevisiae*. *Cell* **89**, 381–391 (1997).
- Dernburg, A. F. *et al.* Meiotic recombination in *C. elegans* initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell* **94**, 387–398 (1998).
- Francis, R., Maine, E. & Schedl, T. Analysis of the multiple roles of *gld-1* in germline development: interactions with the sex determination cascade and the *glp-1* signalling pathway. *Genetics* **139**, 607–630 (1995).
- Jones, A. R., Francis, R. & Schedl, T. GLD-1, a cytoplasmic protein essential for oocyte differentiation, shows stage- and sex-specific expression during *Caenorhabditis elegans* germline development. *Dev. Biol.* **180**, 165–183 (1996).
- SenGupta, D. J. *et al.* A three-hybrid system to detect RNA-protein interactions *in vivo*. *Proc. Natl Acad. Sci. USA* **93**, 8496–8501 (1996).
- Bernstein, D. S., Buter, N., Stumpf, C. & Wickens, M. Analyzing mRNA-protein complexes using a yeast three-hybrid system: methods and applications. *Methods* **26**(3), 123–141 (2002).
- Jones, A. R. & Schedl, T. Mutations in *gld-1*, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. *Genes Dev.* **9**, 1491–1504 (1995).
- Jan, E., Motzny, C. K., Graves, L. E. & Goodwin, E. B. The STAR protein, GLD-1, is a translational regulator of sexual identity in *Caenorhabditis elegans*. *EMBO J.* **18**, 258–269 (1999).
- Lee, M. H. & Schedl, T. Identification of *in vivo* mRNA targets of GLD-1, a maxi-KH motif containing protein required for *C. elegans* germ cell development. *Genes Dev.* **15**, 2408–2420 (2001).
- Xu, L., Paulsen, J., Yoo, Y., Goodwin, E. B. & Strome, S. *Caenorhabditis elegans* MES-3 is a target of GLD-1 and functions epigenetically in germline development. *Genetics* **159**, 1007–1017 (2001).
- Kraemer, B. *et al.* NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in *Caenorhabditis elegans*. *Curr. Biol.* **9**, 1009–1018 (1999).
- Ward, S., Roberts, T. M., Strome, S., Pavalko, F. M. & Hogan, E. Monoclonal antibodies that recognize a polypeptide antigenic determinant shared by multiple *Caenorhabditis elegans* sperm-specific proteins. *J. Cell Biol.* **102**, 1778–1786 (1986).
- Crittenden, S. L., Troemel, E. R., Evans, T. C. & Kimble, J. GLP-1 is localized to the mitotic region of the *C. elegans* germ line. *Development* **120**, 2901–2911 (1994).
- Hendzel, M. J. *et al.* Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106**, 348–360 (1997).
- Wharton, R. P. & Struhl, G. RNA regulatory elements mediate control of *Drosophila* body pattern by the posterior morphogen *nanos*. *Cell* **67**, 955–967 (1991).
- Tadauchi, T., Matsumoto, K., Herskowitz, I. & Irie, K. Post-transcriptional regulation through the HO 3'-UTR by Mpt5, a yeast homolog of Pumilio and FBF. *EMBO J.* **20**, 552–561 (2001).

Acknowledgements

We thank R. Reijo Pera for sharing unpublished observations; and members of the Kimble and Wickens laboratories, T. Schedl and E. Goodwin for comments on the manuscript. J.K. is an investigator of the Howard Hughes Medical Institute. M.W. is supported by the National Institutes of Health.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.K. (e-mail: jekimble@facstaff.wisc.edu).

Kremen proteins are Dickkopf receptors that regulate Wnt/ β -catenin signalling

Bingyu Mao*, Wei Wu*, Gary Davidson*, Joachim Marhold†, Mingfa Li†, Bernard M. Mechler†, Hajo Delius‡, Dana Hoppe*, Peter Stannek*, Carmen Walter*, Andrei Glinka* & Christof Niehrs*

* Molecular Embryology Division, † Developmental Genetics Division, and ‡ Applied Tumor Virology Division, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

The Wnt family of secreted glycoproteins mediate cell–cell interactions during cell growth and differentiation in both embryos and adults^{1,2}. Canonical Wnt signalling by way of the β -catenin pathway is transduced by two receptor families. Frizzled proteins and lipoprotein-receptor-related proteins 5 and 6 (LRP5/6) bind Wnts and transmit their signal by stabilizing intracellular β -catenin^{3–6}. Wnt/ β -catenin signalling is inhibited by the secreted protein Dickkopf1 (Dkk1), a member of a multi-gene family, which induces head formation in amphibian embryos⁷. Dkk1 has been shown to inhibit Wnt signalling by binding to and antagonizing LRP5/6^{8–10}. Here we show that the transmembrane proteins Kremen1 and Kremen2 are high-affinity Dkk1 receptors that functionally cooperate with Dkk1 to block Wnt/ β -catenin signalling. Kremen2 forms a ternary complex with Dkk1 and LRP6, and induces rapid endocytosis and removal of the Wnt receptor LRP6 from the plasma membrane. The results indicate that Kremen1 and Kremen2 are components of a membrane complex modulating canonical Wnt signalling through LRP6 in vertebrates.

We screened pools of clones from an embryonic mouse complementary DNA library by expression in 293T cells and selecting pools that confer binding of recombinant Dkk1–alkaline phosphatase fusion protein (Dkk1–AP¹⁰). Screening a total of 4×10^5 clones, we identified two complementary DNAs that mediated Dkk1–AP binding. Both cDNAs were independent isolates of *kremen2* (*krm2*), a relative (35.3% amino-acid sequence identity) of a gene that encodes a putative type I transmembrane protein *kremen1* (*krm1*) of unknown function and without invertebrate orthologues¹¹. Both Krm proteins contain an extracellular kringle-, WSC- and CUB domain, and a cytoplasmic domain without conserved motifs (see below). Both Dkk1–AP and Dkk2–AP¹⁰ bind with high affinity to 293T cells transfected with *krm1* or *-2* (Fig. 1a–f). Neither AP alone (Fig. 1a, g) nor Dkk3–AP¹⁰ (Fig. 1b, d) nor myc–Xwnt8 protein¹² (not shown) bind *krm*-transfected cells. Scatchard analysis reveals bimodal binding to Krm1, indicating a low- and a high-affinity binding site for Dkk1–AP and Dkk2–AP (Fig. 1e, f). Dkk1–AP binding to *krm2*-transfected cells is competed by recombinant Dkk1 and Dkk2 but not Dkk3 (Fig. 1g). Binding of Dkk1 to Kremen is direct, as shown by a cell surface crosslinking experiment followed by co-immunoprecipitation. Dkk1–AP is specifically immunoprecipitated by flag-tagged Krm2 but not by Krm2 lacking the WSC domain (Fig. 1h). We conclude that Krm1 and *-2* are high-affinity receptors for Dkk1 and *-2*.

To determine if Dkk1 and Kremen interact functionally during Wnt signalling, we used 293T cells transfected with the luciferase reporter construct TOPFLASH, carrying TCF-binding sites, which is directly activated by the TCF/ β -catenin complex¹³. Wnt/*fz* cotransfection in 293T cells induces the TOPFLASH reporter (Fig. 2a, bar 2) and neither *krm1* nor *-2* significantly inhibit this response (bars 4, 5). However, cotransfection of *krm1* or *-2* with a limiting amount of *dkk1* leads to strong synergistic inhibition of Wnt signalling (bars 6, 7). As shown previously, LRP6 can out-

titrate the ability of *dkk1* to inhibit Wnt signalling^{8,10} (bars 8, 9), but cotransfection of *dkk1* with *krm1* or *-2* overcomes this limitation (bars 12, 13).

Drosophila has no *dkk* nor *krm* but an LRP6 homologue, *arrow*, which functions in Wnt signalling⁵. To determine if they could inhibit Wnt signalling in the fly, we expressed *Xenopus dkk1* and mouse *krm2* as heterologous transgenes in *Drosophila* and used the GAL4/UAS system¹⁴ with a *scalloped* (*sd*)-GAL4 driver to direct their expression to the wing disc. Development of the wing critically depends on Wnt signalling, and interference with *wingless* or components of the Wnt pathway characteristically results in loss of wing structures¹⁵. Even though Dkk1 protein is produced in transgenic flies (Fig. 2b), it does not affect wing development by itself (Fig. 2c). However, coexpression of *dkk1* and *krm2* results in almost complete loss of wings, whereas expression of *krm2* alone has no effect (Fig. 2d, e). These results indicate that Krm is required for inhibition of Wnt signalling by Dkk1, presumably by interacting with *arrow*. Indeed, Dkk1 binds to and functionally interacts with *Drosophila arrow* transfected in 293T cells (data not shown). Furthermore, inhibition of *krm1* and *krm2* by antisense Morpholino oligonucleotides reveals that they are required for Wnt inhibition during embryonic head formation and interact with *dkk1* in *Xenopus* (G.D., B.M. and C.N., manuscript in preparation).

Next we analysed which domains of Krm2 are required for the physical and functional interaction with Dkk1 by testing deletion constructs expressed in 293T cells (Fig. 3a, b). Deletion of any of the extracellular domains (ECD) abolishes interaction with Dkk1 in binding assays (Fig. 3c, e) and Wnt reporter assays (Fig. 3d, e). In contrast, following deletion of the intracellular domain (Δ C) significant binding and Wnt inhibition are retained (Fig. 3c–e).

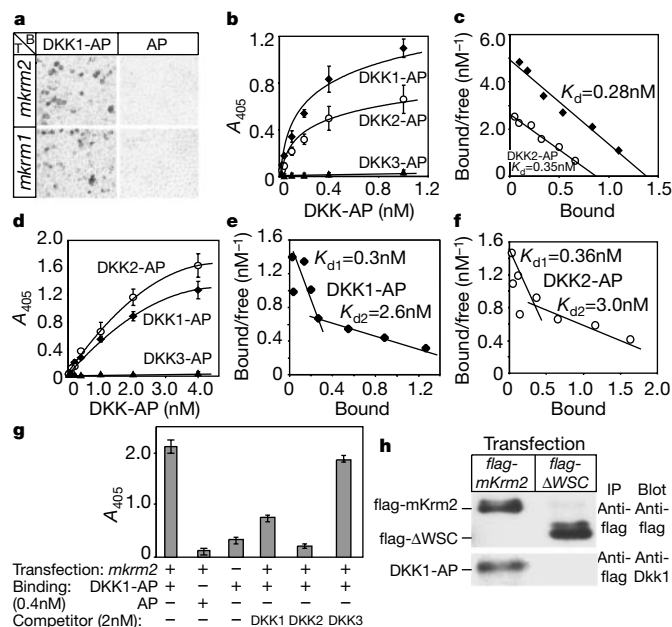


Figure 1 Kremens are high-affinity Dkk receptors. **a**, Dkk1–AP but not soluble AP binds to cells transfected with murine *krm1* or *krm2*. 293T cells were transfected (T) with *mkrm1* or *mkrm2* as indicated, incubated with recombinant Dkk1–AP or AP (B) and stained for bound AP activity. **b, c**, Binding curves and Scatchard analysis of Dkk–AP fusion proteins binding to *mkrm2* transfected cells. **d**, Binding curves for Dkk–APs binding to *mkrm1* transfected cells. Scatchard analysis for Dkk1–AP (**e**) and Dkk2–AP (**f**) binding to *mkrm1* transfected cells. Note two binding sites in *mKrm1* for Dkks. Dissociation constants (K_d) are indicated. **g**, Binding of Dkk1–AP to *mkrm2* transfected cells can be specifically inhibited by Dkk1 and mDkk2 but not mDkk3 protein. **h**, Dkk1–AP can be specifically crosslinked on cell surface to flag–mKrm2, but not to flag–mKrm2– Δ WSC (see Fig. 3a). A_{405} , absorbance at 405 nm.

Membrane attachment of the Krm2 ECD is critical for its function, as a secreted version (Δ TMC) mediates no Wnt inhibition (Fig. 3d) whereas ECD membrane attachment via a GPI-anchor (myc-GPI) confers binding and Wnt inhibition (Fig. 3c–e). The results indicate that all extracellular domains of Krm2 are required for functional interaction with Dkk1, but that its cytoplasmic domain is of minor importance for its ability to promote inhibition of Wnt signalling.

To test if Krm2 interacts with LRP6, we produced the extracellular domain of Krm2 as V5-tagged, secreted recombinant protein (mKrm2 Δ TMC) and tested its binding to LRP6-transfected 293T cells. mKrm2 Δ TMC protein binds effectively to LRP6 transfected cells only when Dkk1 protein is added (Fig. 4a, b). No binding is seen when Dkk1 is replaced by Dkk3 or when LRP6 Δ E1–4, lacking the EGF repeats required for Dkk1 binding¹⁰, is transfected (Fig. 4c, d). The interaction between LRP6 and Krm2 is also observed when partners are reversed: The secreted recombinant extracellular domain of LRP6 (LRP6N-myc⁴) binds to *krm2* transfected cells only in the presence of Dkk1 (Fig. 4e, f). This interaction does not require the cytoplasmic tail of Krm2 but an intact ECD, as deletion of the WSC domain abolishes binding (Fig. 4g, h).

To test if Krm and LRP6 also interact in live cells, we used bioluminescence resonance energy transfer (BRET) assays^{16,17}. Renilla luciferase was fused to the cytoplasmic tail of LRP6, and yellow fluorescent protein (YFP) was fused to Krm2. If interactions between the two fusion proteins bring the luciferase and YFP in

close contact, resonance energy transfer occurs, changing the colour of the bioluminescent emission, which is measured ratiometrically. An increase in the BRET ratio is thus indicative of molecular interaction between the tagged proteins. Coexpression in 293T cells of luciferase-tagged LRP6 with YFP-tagged-*krm1* or -2 results in a basal BRET signal which is enhanced by *dkk1* and -2 but not by *dkk3* (Fig. 4i). This BRET signal increase is abolished by deleting the kringle domain in Krm2 (Fig. 4i, bars 9 and 10), which is required for Dkk interaction (Fig. 3c, e). In a real-time assay the BRET signal of the two receptors remains stable over minutes in controls and after addition of recombinant Dkk3 protein (Fig. 4j). However, Dkk1 protein leads to an increase of the BRET signal within 5 min of addition. Taken together, these results show that Dkk1 triggers formation of a ternary receptor complex involving Krm2 and LRP6. Dkk1 may bridge these receptors directly.

To analyse the consequence of ternary complex formation, we investigated the fate of LRP6 using a LRP6–GFP fusion in transfected 293T cells. LRP6–GFP is localized at the plasma membrane, as well as in intracellular vesicles in untreated cells (not shown) or in cells which are treated at 4 °C or 37 °C with Dkk1 protein (Fig. 5a, middle, 5b, middle). However, when cotransfected with *krm2*, LRP6–GFP plasma membrane localization is no longer observed after 30 min treatment with Dkk1 at 37 °C. Instead, LRP6–GFP is seen only in intracellular vesicles, where it colocalizes with Dkk1 (Fig. 5c), indicating that both proteins are co-internalized. When endocytosis is blocked by incubation at 4 °C, LRP6–GFP and Dkk1 remain at the cell surface (Fig. 5d). These results were confirmed by measuring plasma membrane levels of LRP6 using cell-surface biotinylation. Following cotransfection with murine *krm2* and addition of Dkk1 protein, cell-surface LRP6 levels are strongly

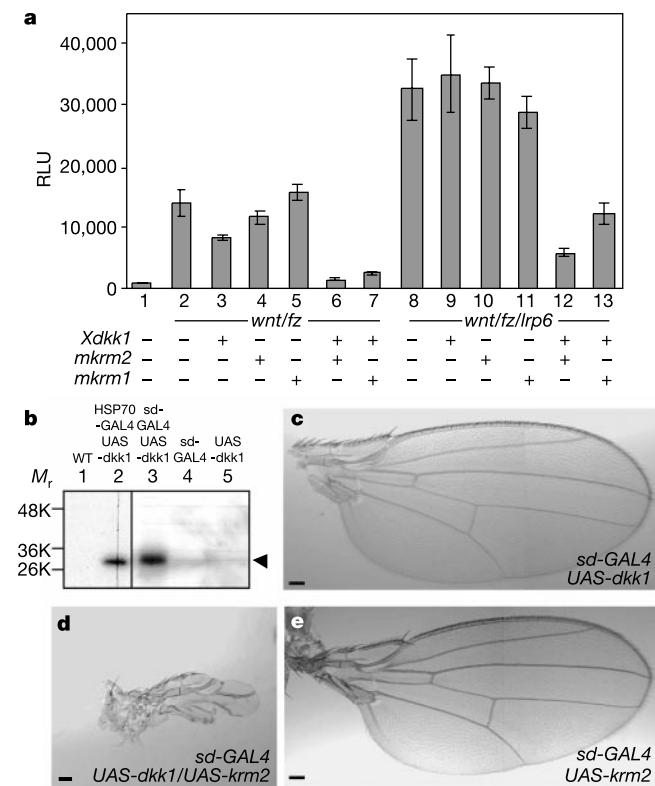


Figure 2 Kremen co-operate with Dkk1 in Wnt/LRP6 signalling inhibition. **a**, Luciferase Wnt reporter assays with the genes indicated. DNA amounts used were (per well): *Xdtk1*, 1 ng (bars 2–7), 3 ng (bars 8–13); *mkrm1*, -2, 1 ng (bars 2–7), 3 ng (bars 8–13); *mwnt1*, 8 ng; *mfz8*, 1 ng; *LRP6*, 3 ng. **b–e**, Heterologous co-expression of *dkk1* and *krm2* leads to wingless phenotype in *Drosophila*. **b**, Western analysis of protein extracts from wild-type (WT) larvae (lane 1), third-instar larvae carrying *UAS-dkk1/hsp70-GAL4* (lane 2), *UAS-dkk1/sd-GAL4* (lane 3), *sd-GAL4* (lane 4) and *UAS-dkk1* (lane 5). Ectopic Dkk1 protein is indicated by the arrowhead. **c–e**, Adult wings from, **c**, *sd-GAL4/+; UAS-dkk1/+* and **d**, *sd-GAL4/+; UAS-dkk1/+; UAS-krm2* **e**, *sd-GAL4/+; UAS-krm2* animals (scale bar, 80 μ m). RLU, relative light units.

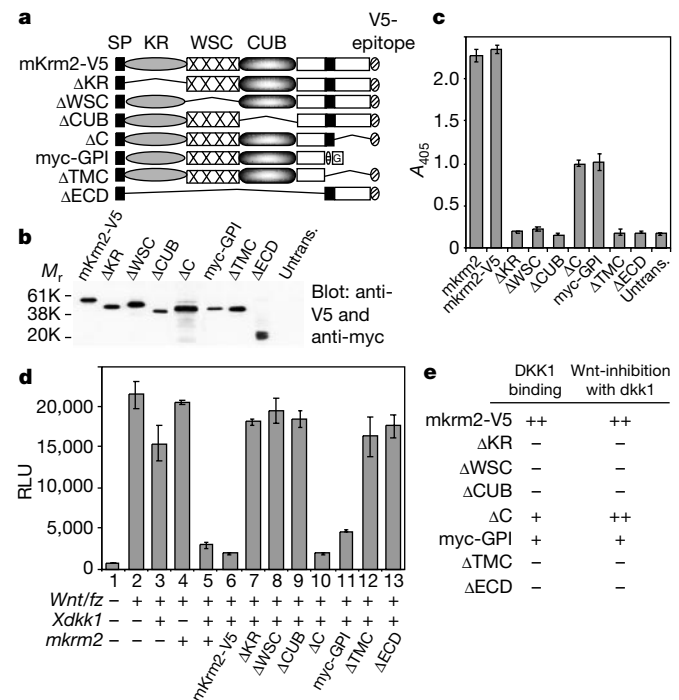


Figure 3 Deletion analysis of Kremen. **a**, Schematic drawing of the *mkrm2* deletion constructs used in this study. SP, signal peptide; KR, kringle domain; TM, transmembrane domain. **b**, Western blot showing the expression of the deletion constructs in transfected 293T cells. **c**, Dkk1-AP binding to 293T cells transfected with *mkrm2* deletion constructs. **d**, Luciferase Wnt reporter assays with the genes indicated. DNA amounts used for transfection: *mwnt1*, 8 ng; *mfz8*, 1 ng; *Xdtk1*, 1 ng; all *mkrm* constructs were used at 2 ng per well. **e**, Summary of the binding and Wnt inhibition of the *mkrm2* deletion constructs shown in **c** and **d**.

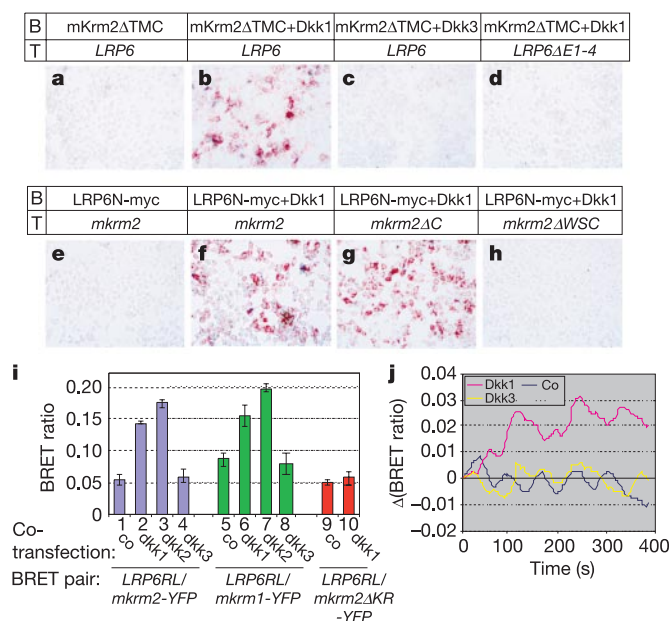


Figure 4 Kremens, Dkk1 and LRP6 form a ternary complex. **a–h**, 293T cells were transfected (T) with the indicated gene. Binding (B) was done with V5-tagged recombinant mKrm2ΔTMC or LRP6N-myc proteins in the presence or absence of Dkk proteins and followed by immunohistochemistry. **i**, 293T cells were transfected with the indicated genes and analysed by BRET assay. Co, control vector. **j**, Real-time BRET assay of 293T cells transfected with LRP6-RL and mkrm1-YFP. At $t = 0$, recombinant Dkk1 or Dkk3 proteins were added and the ensuing change of the BRET ratio Δ (BRET ratio) is plotted. Co, control. This experiment was repeated four times independently with similar results.

reduced, indicative of receptor clearance from the plasma membrane (Fig. 5e, lane 4). Similar to ternary complex formation, receptor clearance is a rapid process and is detectable within 5 min of Dkk1 application (Fig. 5f). A slight decrease in total LRP6 level is also observable (anti-flag lanes in Fig. 5e, f), which may indicate that the protein is degraded following internalization.

Our results suggest a model whereby Dkk1 inhibits Wnt signalling by acting in concert with its receptor Kremen to form a ternary complex with LRP6, which is rapidly endocytosed. This eliminates the Wnt receptor from the plasma membrane, thus preventing Wnt–LRP6 interaction (Fig. 5g).

Methods

Expression cloning of murine Kremen2 and constructs

A mouse 13.5-d-embryo cDNA library in the expression vector pCMV-SPORT2 (Gibco BRL) was used to prepare pools of about 250 colonies, and plasmid DNA from each pool was transiently transfected into 293T cells in 24-well plates using FuGENE 6 (Roche). After 48 h, cells were incubated with medium containing 1 nM Dkk1–AP fusion protein¹⁰, and processed for AP histochemistry¹⁰. From 1,500 pools, 2 positive pools were identified and single clones were isolated by sib selection. Sequencing analysis showed that they represent independent isolates of *mkremen2*. A full length *mkremen1*¹¹ clone was isolated from the same library by polymerase chain reaction, PCR. The open reading frame of *mkremen1* and -2 was cloned into pCS2 +¹⁸ to generate pCS2-mkrm1 and -2. pCS-flag-mkrm2 was constructed by inserting a flag epitope after the signal peptide and was used as template to generate the pCS-flag-mkrm2ΔWSC by PCR. pCS-mkrm2-V5 was generated by inserting mkrm2 into a vector containing a C-terminal V5-tag. The mkrm2 deletion series was made by PCR using pCS-mkrm2-V5 as template. A PRK5-myc-GPI vector was used to generate the mkrm2-myc-GPI construct. To create the pLRP6-RL and LRP6-GFP construct, fragments encoding Renilla luciferase (RL) or GFP was amplified by PCR and fused C-terminally to LRP6⁴. pCS-mkrm1,2-YFP were made by fusing EYFP (Clontech) C-terminally.

Cell surface binding, biotinylation, crosslinking, immunoprecipitations

Transfections to obtain recombinant protein conditioned media, cell surface binding and competition analysis were done as described¹⁰. For the crosslinking experiments (Fig. 1h), cells in 10-cm dishes were transfected with pCS-flag-mkrm2 or a control pCS-flag-mkrm2-ΔWSC, and after 48 h cells were incubated with medium containing Dkk1–AP (2 nM) for 1 h on ice, washed twice with PBS, and crosslinked with dithiobis[succinimidyl

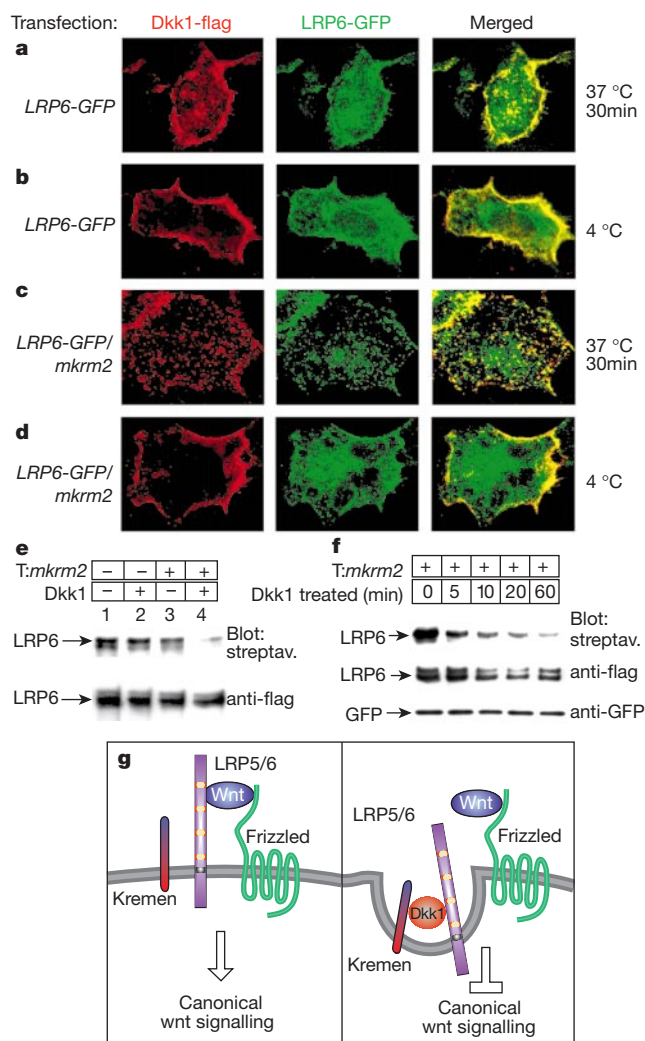


Figure 5 Dkk1 and Kremen trigger removal of LRP6 from the cell surface. **a–d**, Left panels, confocal images showing the localization of Dkk1-flag (red) bound to 293T cells that were transfected with LRP6–GFP with or without *mkrm2*. Cells were incubated at 4 °C with Dkk1-flag conditioned medium, washed, and then incubated for 30 min at 37 °C or at 4 °C as indicated. The distribution of LRP6–GFP in the same cells is shown in green (**a–d**, middle panels), and the merged pictures are shown in **a–d**, right panels. **e**, Depletion of cell surface LRP6 by Dkk1/Kremen. 293T cells were transfected with flag-LRP6 and with or without *krm2* as indicated. After 2 d, cells were treated with Dkk1 conditioned medium as indicated for 1 h at 37 °C and washed. After cell surface biotinylation, cell lysates were immunoprecipitated with anti-flag beads and the blots were probed with streptavidin-HRP (upper panel) to detect cell surface LRP6 levels and anti-flag (lower panel) to detect total LRP6. **f**, 293T cells were transfected with flag-LRP6, *mkrm2* and GFP (transfection/lysis control) and treated at 37 °C with Dkk1 conditioned medium for the times indicated. Subsequently, cells were surface-biotinylation and analysed as in **e**. The lower panel shows western blot detection of GFP from total lysates (transfection/lysis control). **g**, A model for Dkk1/Kremen action. LRP5/6 act as co-receptors for the canonical Wnt pathway (left). Dkk1, LRP5/6 and Kremen form a ternary complex that triggers rapid internalization and depletion of cell-surface LRP5/6 and thus blockage of canonical Wnt signalling (right).

propionate] (DSP) (Pierce, 0.4 mM) for 30 min on ice. Cell lysates were then subjected to immunoprecipitation with anti-flag beads (Sigma). The immunoprecipitates were prepared for western blots after de-crosslinking with dithiothreitol. For Fig. 4a–h, transfected 293T cells in 24-well plates were placed on ice for 10 min, washed with ice-cold DMEM/10%FCS, and incubated with the indicated proteins for 2 h. Then cells were washed with Hank's saline and fixed for 10 min with 4% paraformaldehyde/PBS for immunostaining. Cell surface biotinylation was carried out on transfected cells as described in Fig. 5e or cotransfected with *mzf8* + *mWnt1* (Fig. 5f) using 0.5 mg ml^{−1} sulpho-NHS-LC-biotin (Pierce) according to the manufacturer. Flag-LRP6 was

immunoprecipitated with anti-flag beads, eluted with flag peptide and analysed by western blot with anti-flag (total LRP6) or HRP-streptavidin (cell surface LRP6).

Luciferase reporter and BRET assays

Luciferase Wnt reporter assays in 293T cells were done in 96-well plates at least in triplicate as described¹⁹. Luciferase activity was normalized against Renilla activity. For BRET assays, 293T cells in 24-well plate were transfected with 200 ng pCS-LRP6-RL and pCS-mkrm-YFP constructs with or without co-transfection of 100 ng *dkks*. 24 hours post-transfection, cells were collected in Hank's buffer (with Ca^{2+} and Mg^{2+}) and transferred in 96-well luminometer plates. Coelenterazine h (Molecular Probe) was added to a final concentration of 5 μM , and light emission (*E*) was measured at 470 nm and 530 nm sequentially using an Ascent FL luminometer (Labsystems). Data are represented as BRET ratio, which is defined as $(E_{530}/E_{470})_{\text{sample}} - (E_{530}/E_{470})_{\text{LRP6-RL only}}$. Expression of all constructs was controlled by luminescence or fluorescence. For real time BRET assay, cells were grown and transfected with 50 ng pCS-LRP6-RL and 20 ng pCS-mkrm1-YFP directly in 96-well luminometer plates, and cell growth and transfection were controlled in parallel in a regular 96-well plate. 30 hours post-transfection, medium was removed and Coelenterazine h (10 μM in Hank's buffer) was added and incubated for 1 min before Dkk proteins (5 nM final) were added. Light emission at 470 nm and 530 nm was measured continuously with 1 s integration per data point at 37 °C. Data were collected and processed with Microsoft Excel by averaging ten data points, and are represented as the absolute changes of the BRET ratios from time $t = 0$.

Drosophila transgenesis

Oregon R flies carrying *UAS-krm2* and *UAS-dkk1* were generated by *P*-element transformation of P[ry⁺; $\Delta 2$ -3]99B flies with the pUAST-Vector containing cDNA insert of pCS2-Xdkk1-flag and pCS2-mkrm2-V5, and bred with *sd-GAL4*-(P[GAL4]sd^{SG29.1}) and *hsp70-GAL4* (Bloomington stock 1799) flies. Protein extracts were prepared from third-instar larvae carrying the *UAS-dkk1* transgene driven by either by *sd-GAL4* or *hsp70-GAL4*. Proteins were separated using 10% SDS-polyacrylamide gel electrophoresis (PAGE), and Dkk1 protein was detected with anti-15 Dkk1 antibody⁷.

Immunofluorescence microscopy

293T cells on cover slips were transfected with LRP6-GFP (300 ng per well) or LRP6-GFP/mkrm2 (300/100 ng per well). 30 hours after transfection, the cells were incubated with Dkk1-flag conditioned medium for 2 h at 4 °C, washed with cold PBS and changed to fresh medium, then one group was kept at 4 °C as control while the other group was incubated at 37 °C for 30 min. Cells were fixed, permeabilized and proceeded for staining with anti-flag antibody and confocal microscopy.

Received 15 February; accepted 11 April 2002; doi:10.1038/nature756.
Published online 26 May 2002.

- Wodarz, A. & Nusse, R. Mechanisms of Wnt signalling in development. *Annu. Rev. Cell Dev. Biol.* **14**, 59–88 (1998).
- Miller, J. R., Hocking, A. M., Brown, J. D. & Moon, R. T. Mechanism and function of signal transduction by the Wnt/ β -catenin and Wnt/ Ca^{2+} pathways. *Oncogene* **18**, 7860–7872 (1999).
- Bhanot, P. *et al.* A new member of the frizzled family from *Drosophila* functions as a Wingless receptor. *Nature* **382**, 225–230 (1996).
- Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530–535 (2000).
- Wehrli, M. *et al.* *arrow* encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**, 527–530 (2000).
- Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J. & Skarnes, W. C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535–538 (2000).
- Glinka, A. *et al.* Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. *Nature* **391**, 357–362 (1998).
- Bafico, A., Liu, G., Yaniv, A., Gazit, A. & Aaronson, S. A. Novel mechanism of Wnt signalling inhibition mediated by Dickkopf-1 interaction with LRP6/Arrow. *Nature Cell Biol.* **3**, 683–686 (2001).
- Semenov, M. V. *et al.* Head inducer Dickkopf-1 is a ligand for Wnt coreceptor LRP6. *Curr. Biol.* **11**, 951–961 (2001).
- Mao, B. *et al.* LDL-receptor-related protein 6 is a receptor for Dickkopf proteins. *Nature* **411**, 321–325 (2001).
- Nakamura, T., Aoki, S., Kitajima, K., Takahashi, T. & Matsumoto, K. Molecular cloning and characterization of Kremen, a novel kringle-containing transmembrane protein. *Biochim. Biophys. Acta* **1518**, 63–72 (2001).
- Hsieh, J. C., Rattner, A., Smallwood, P. M. & Nathans, J. Biochemical characterization of Wnt-frizzled interactions using a soluble, biologically active vertebrate Wnt protein. *Proc. Natl Acad. Sci. USA* **96**, 3546–3551 (1999).
- Korinek, V. *et al.* Constitutive transcriptional activation by a β -catenin-Tcf complexes in APC/ colon carcinoma. *Science* **275**, 1784–1787 (1997).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
- Couso, J. P., Bishop, S. A. & Martinez Arias, A. The wingless signalling pathway and the patterning of the wing margin in *Drosophila*. *Development* **120**, 621–636 (1994).
- Angers, S. *et al.* Detection of β 2-adrenergic receptor dimerization in living cells using bioluminescence resonance energy transfer (BRET). *Proc. Natl Acad. Sci. USA* **97**, 3684–3689 (2000).
- Xu, Y., Piston, D. W. & Johnson, C. H. A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. *Proc. Natl Acad. Sci. USA* **96**, 151–156 (1999).
- Rupp, R. A., Snider, L. & Weintraub, H. *Xenopus* embryos regulate the nuclear localization of XMyoD. *Genes Dev.* **8**, 1311–1323 (1994).
- Wu, W., Glinka, A., Delius, H. & Niehrs, C. Mutual antagonism between dickkopf1 and -2 regulates Wnt/ β -catenin signalling. *Curr. Biol.* **10**, 1611–1614 (2000).

Acknowledgements

We thank X. He for providing reagents and advice, and J. Nathans and P. Ramulu for pRKdkk1-AP and other plasmids. Other materials were provided by H. Clevers, S. Cohen, S. Di Nardo, J. Flanagan, R. Moon and R. Nusse. We thank S. Cohen for advice with fly work and fly stocks, H. Spring for confocal microscopy and U. Hebling for sequencing. This work was supported by the Deutsche Forschungsgemeinschaft.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.N.

(e-mail: Niehrs@DKFZ-Heidelberg.DE). The sequence of murine *kremen2* has been deposited with GenBank (accession number AJ457192).

New components of the spliced leader RNP required for nematode *trans*-splicing

John A. Denker, David M. Zuckerman, Patricia A. Maroney & Timothy W. Nilsen

Center for RNA Molecular Biology, Case Western Reserve University School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA

Pre-messenger-RNA maturation in nematodes and in several other lower eukaryotic phyla involves spliced leader (SL) addition *trans*-splicing^{1,2}. In this unusual RNA processing reaction, a short common 5' exon, the SL, is affixed to the 5'-most exon of multiple pre-mRNAs. The nematode SL is derived from a *trans*-splicing-specific ~100-nucleotide RNA (SL RNA) that bears striking similarities to the *cis*-spliceosomal U small nuclear RNAs U1, U2, U4 and U5 (refs 3, 4); for example, the SL RNA functions only if it is assembled into an Sm small nuclear ribonucleoprotein (snRNP)⁵. Here we have purified and characterized the SL RNP and show that it contains two proteins (relative molecular masses 175,000 and 30,000 (M_r 175K and 30K)) in addition to core Sm proteins. Immunodepletion and reconstitution with recombinant proteins demonstrates that both proteins are essential for SL *trans*-splicing; however, neither protein is required either for conventional *cis*-splicing or for bimolecular (*trans*-) splicing of fragmented *cis* constructs. The M_r 175K and 30K SL RNP proteins are the first factors identified that are involved uniquely in SL *trans*-splicing. Several lines of evidence indicate that the SL RNP proteins function by participating in a *trans*-splicing specific network of protein-protein interactions analogous to the U1 snRNP-SF1/BBP-U2AF complex that comprises the cross-intron bridge in *cis*-splicing.

Previously, we identified two candidate SL RNP-specific proteins of M_r ~30K and 175K whose association with the *Ascaris* SL RNP was correlated with function in *trans*-splicing⁶. To characterize these proteins further, the SL RNP was purified in amounts sufficient for protein sequencing. This analysis yielded several unique peptides from both the M_r 30K and 175K proteins. Using degenerate oligonucleotides and a variety of polymerase chain reaction (PCR) approaches, we obtained what seemed to be full-length complementary DNA clones encoding both proteins; both cDNAs contained all of the expected peptides and each had a single long open reading frame that began with a methionine codon and ended with multiple stop codons (see Supplementary Information). Other than presumptive orthologues in *Caenorhabditis elegans* (see Supplementary Information), computer searches did not reveal any proteins with significant similarity. Furthermore, neither protein

contained any characterized motif suggestive of function. Although the larger protein had an M_r of 175K, conceptual translation of its corresponding cDNA clone predicted a protein of only ~95K. To ensure that we had indeed obtained full-length clones, synthetic mRNAs corresponding to the cDNA clone were translated in rabbit reticulocyte lysate. The translation product migrated at M_r ~175K (data not shown). We do not yet know why the protein migrates

aberrantly on denaturing gels.

To study the proteins further, polyclonal antisera were raised against bacterially produced fragments of each; M_r 30K-specific antisera recognized a single *Ascaris* protein of appropriate size on Western blots, as did the 175K-specific antisera (see Fig. 1b). To determine whether either or both proteins were associated with one or more RNAs, immunoprecipitation analyses were performed. Strikingly, both antisera precipitated a single RNA, the SL RNA, indicating that we had indeed cloned proteins associated with the SL RNP and that the proteins were SL RNP-specific (Fig. 1a). We then

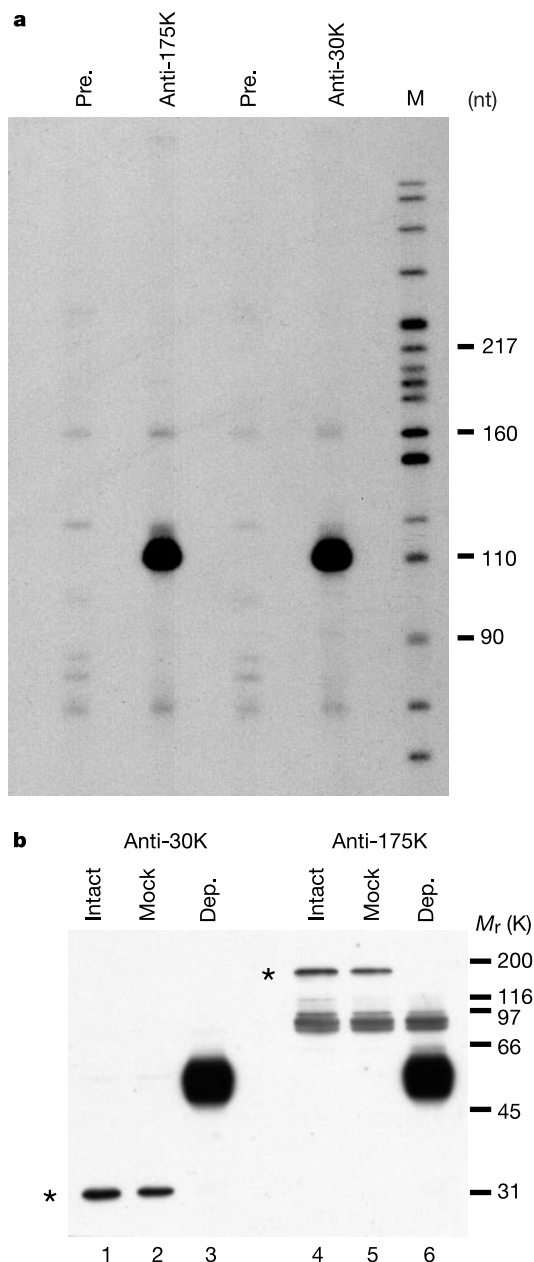


Figure 1 The M_r 30K and 175K proteins are SL RNP-specific. **a**, RNAs present in immunoprecipitates of *Ascaris* whole-cell extract were labelled and fractionated on denaturing gels (see Methods); the antisera used are indicated. The single-labelled RNA present in the anti-175K and anti-30K lanes was identified as SL RNA by digestion with RNase H in the presence of SL-specific oligodeoxynucleotides (data not shown). Pre., preimmune serum; M, DNA markers. **b**, Immunoblot probed with the indicated antisera; lanes 1 and 4, intact extract; lanes 2 and 5, mock-depleted extract; lanes 3 and 6, depleted extract (Dep.) (see Methods). The asterisks denote the 30K and 175K proteins. In lanes 3 and 6, the dark signal arises from residual antibody after immunodepletion. In lanes 4–6, the signal of M_r ~80K is unrelated to the 175K protein and is observed with preimmune sera.

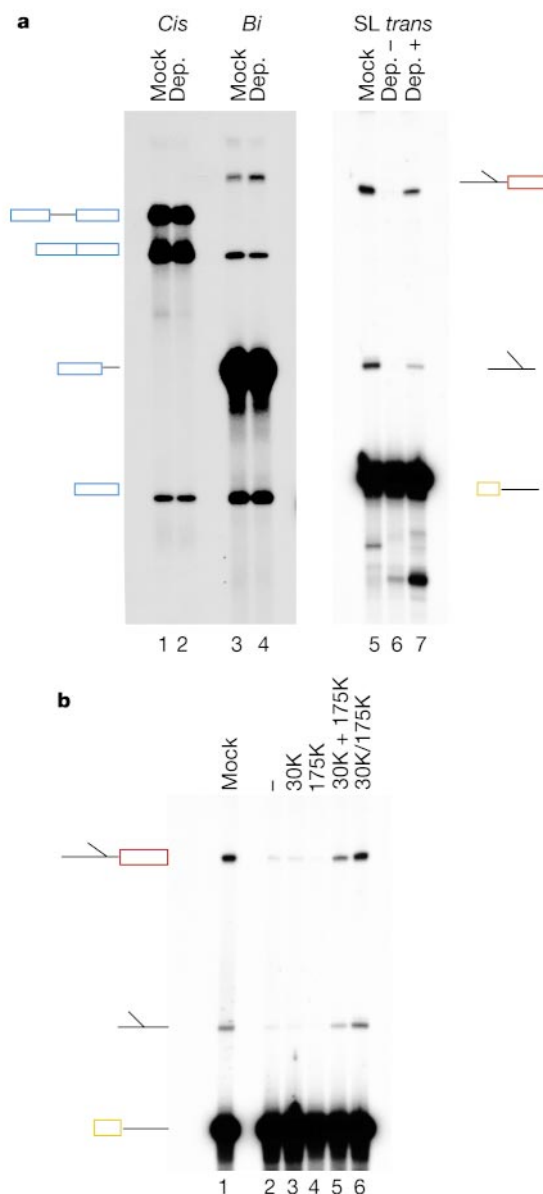


Figure 2 Both SL-RNP-specific proteins are required for *trans*-splicing. **a**, *Cis*, bimolecular (Bi) and SL *trans*-splicing were assayed in mock-depleted extract or extract sequentially depleted of the M_r 30K and 175K proteins (see Methods). In lane 7, the depleted extract was supplemented with 20 ng of preformed 30K/175K heterodimer. The slowly migrating labelled RNA in lanes 3 and 4 is the Y-branched intermediate generated by the first step of bimolecular splicing; other labelled RNAs are indicated schematically. **b**, Splicing reactions performed as in lanes 5–7 in **a**, except in a different depleted extract. Lane 2 contained no additional protein. Lane 3 was supplemented with 20 ng of 30K protein; lane 4 was supplemented with 20 ng of 175K protein; lane 5 was supplemented with 20 ng each of the 30K and 175K proteins; lane 6 was supplemented with 20 ng of preformed 30K/175K heterodimer.

prepared recombinant proteins for functional analyses. An amino-terminal glutathione *S*-transferase (GST)-tagged version of the M_r 30K protein was well expressed in bacteria and could be purified to apparent homogeneity with glutathione-Sepharose (data not shown). We were unable to express full-length M_r 175K protein

in bacteria but obtained reasonable yields of N-terminal His-tagged protein from baculovirus-infected insect cells. Immunoprecipitation assays indicated that both recombinant proteins were functional for SL RNP assembly; as expected from previous studies⁶, binding of the SL RNP-specific proteins depended on prior association of Sm proteins (data not shown). Additional analyses, including reciprocal pulldown and co-immunoprecipitation, showed that the M_r 175K and 30K proteins interact physically, suggesting that the SL-specific proteins bind to the Sm SL RNP as a preformed heterodimer (see below).

Having established that the M_r 175K and 30K proteins were SL RNP-specific, we examined whether they were necessary for SL RNP function in *trans*-splicing. In extracts immunodepleted for the SL-specific proteins, SL *trans*-splicing with either endogenous or, more importantly, exogenously supplied SL RNA was markedly inhibited (data not shown, and Fig. 2a). Remarkably, in the same extracts, *cis*-splicing activity was not affected (Fig. 2a). As a further control for specificity, we examined whether immunodepleted extracts could catalyse the bimolecular (*trans*-) splicing of a '*cis*' construct in which 5' and 3' exons and their accompanying splicing signals were physically separated (for example, refs. 7, 8). Figure 2a (Bi) shows that the joining of such separated exons was not affected by immunodepletion. This result provides the first direct evidence that SL *trans*-splicing and the *trans*-splicing of '*cis*' constructs in pieces' are mechanistically distinct processes.

To ensure that the inhibition of SL *trans*-splicing observed in immunodepleted extracts was due to the depletion of the SL-specific proteins and not the fortuitous and simultaneous depletion of one or more other factors, reconstitution studies were performed. Supplementation with either the M_r 175K protein or the 30K protein alone did not restore SL *trans*-splicing; however, the addition of both proteins simultaneously or supplementation with preformed 175K/30K heterodimer restored SL *trans*-splicing to levels comparable to those in mock-depleted extracts (Fig. 2a, b). We conclude that both SL-specific proteins are required for SL *trans*-splicing.

Because it seemed possible that the function of the SL-specific proteins was to establish interactions between the SL RNP and other components of the splicing apparatus, we performed yeast two-hybrid analyses to identify potential interacting partners. Clones (1.5×10^6) were screened with the M_r 175K protein as bait and 31 confirmed positives were obtained, all of which corresponded to the 30K protein. When the 30K protein was used as bait, 35 confirmed positives were obtained from 3.0×10^6 clones screened. Of these, 32 corresponded to the 175K protein; intriguingly, the remaining 3 corresponded to the *Ascaris* orthologue of the splicing factor SF1/BBP (Fig. 3a). SF1/BBP specifically binds to the branch-point recognition sequence of the pre-mRNA early in spliceosome assembly, and several lines of evidence indicate that this protein is of central importance in the formation of a network of protein-protein interactions that bridges the two ends of *cis* introns (Fig. 3e, *cis*; and see ref. 9). In this regard, in yeast, SF1/BBP can establish simultaneous contacts with a U1 snRNP-specific protein (Prp40p) bound at the 5' splice site and Mud2p (a U2AF-like protein) bound to the 3' splice site⁹. A similar cross-intron bridge is thought to exist in higher eukaryotes; however, a definitive Prp40p homologue has yet to be identified⁹. The same bridge cannot be formed in SL *trans*-splicing because U1 snRNP does not participate in this reaction¹⁰. Indeed, a fundamental unanswered question in SL *trans*-splicing is how the 5' splice site present in the SL RNA associates efficiently with the 3' splice site present in the target pre-mRNA.

The two hybrid results indicated that the interaction of M_r 30K protein with SF1/BBP could mediate such an association. To investigate this possibility we first confirmed the two-hybrid interaction by reciprocal pulldown analyses (data not shown). We then examined whether the interaction could occur in the context of a fully assembled SL RNP. As shown in Fig. 3b, an immobilized

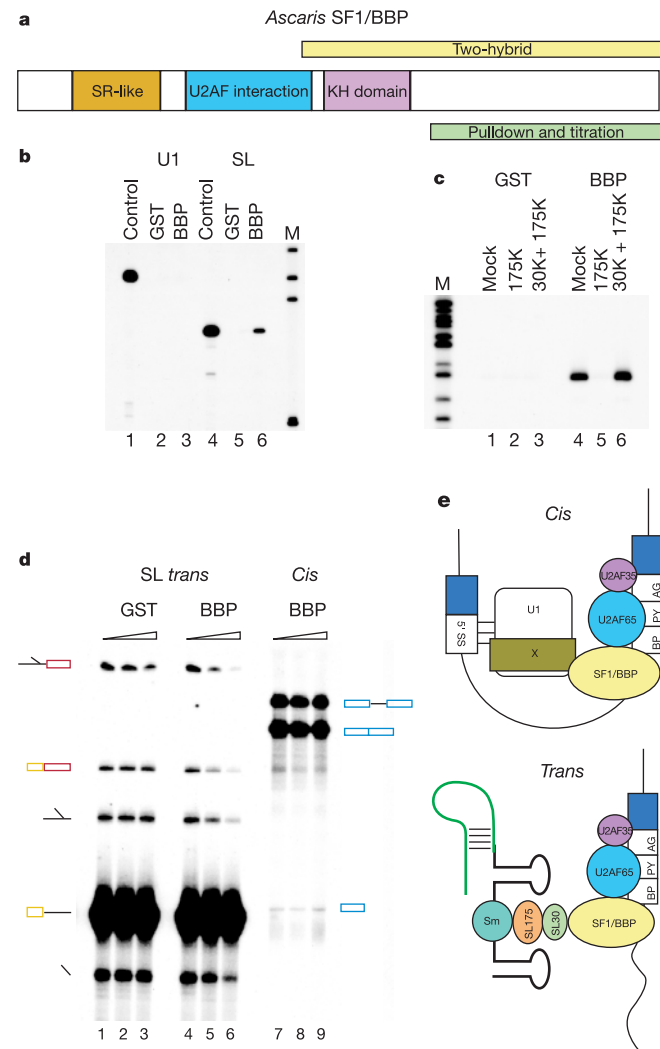


Figure 3 The M_r 30K SL RNP-specific protein interacts functionally with the splicing factor SF1/BBP. **a**, Schematic of *Ascaris* SF1/BBP, which shares 65% identity and 75% similarity with its *C. elegans* homologue; domains are depicted as described previously¹⁷. The portion of the protein recovered by two-hybrid screening is indicated, as is the fragment used for pulldown assays (**b**, **c**) and competition/titration experiments (**d**). **b**, *Ascaris* extract was passed over immobilized GST or GST-BBP (indicated). Bound RNAs were recovered from the reduced glutathione eluate and subjected to primer extension analysis with oligonucleotides complementary to either *Ascaris* U1 small nuclear RNA (lanes 1–3) or SL RNA (lanes 4–6). In lanes 1 and 4, RNA from 1 μ l intact extract was primer-extended. **c**, *Ascaris* extract either mock-depleted (lanes 1 and 4) or depleted of the SL-specific proteins (lanes 2, 3, 5 and 6) (see Methods). The depleted extract was supplemented with either recombinant M_r 175K protein alone (lanes 2 and 4) or both the 175K and 30K proteins (lanes 3 and 6) as in Fig. 2b. Labelled SL RNA was added to each extract and, after assembly, extracts were diluted and passed over columns containing immobilized GST or GST-BBP as indicated. Bound material was recovered from the reduced glutathione eluate and then fractionated on denaturing gels. **d**, SL *trans*-splicing and *cis*-splicing assayed as in Fig. 2b except that the SL RNA was uniformly labelled. Lanes 1–3 were supplemented with 0, 2 and 6 μ g of GST respectively, whereas lanes 4–6 and 7–9 were supplemented with equivalent amounts of GST-BBP. **e**, Schematic representation of *cis* and *trans* splice site bridging complexes. The *cis* model is adapted from ref. 9; for details see the text.

fragment of SF1/BBP specifically selected the SL RNP from unfractionated extract; reconstitution experiments demonstrated that the interaction of the SL RNP with SF1/BBP required the M_r 30K protein (Fig. 3c). We then used a competition approach designed to disrupt the 30K–SF1/BBP interaction in intact extracts. As shown in Fig. 3d, the inclusion of excess polypeptide corresponding to the 30K interaction region of SF1/BBP specifically inhibited SL *trans*-splicing; under these conditions, the polypeptide prevents interaction of the 30K protein with full-length SF1/BBP as assayed by pulldown analysis (data not shown). The same fragment of SF1/BBP had no effect on *cis*-splicing, a result consistent with the fact that this region of SF1/BBP is not required for *cis*-splicing¹¹. Collectively, these analyses indicate that the 30K–SF1/BBP interaction is functionally relevant to SL *trans*-splicing. The striking parallels between our results and those obtained in *cis*-splicing systems strongly suggests that the 30K–SF1/BBP interaction fulfils, in *trans*-splicing, a bridging function analogous to the U1 snRNP–SF1/BBP cross-intron interaction (Fig. 3e).

We have thus identified and characterized two SL RNP intrinsic proteins that are essential for SL *trans*-splicing but not for *cis*-splicing. The apparent absence of similar proteins from human, fly and plant genomes is consistent with the fact that there is no evidence for SL *trans*-splicing in these organisms. It will be of considerable interest to determine whether orthologues of the SL-specific proteins are present in other organisms that do process their mRNAs by *trans*-splicing (for example, *Hydra* and trypanosomes (reviewed in ref. 12)). The presence of such proteins in divergent phyla would indicate that SL *trans*-splicing is an ancestral trait and would thus answer the long-standing debate of whether the process arose once or several times in the course of eukaryotic evolution^{12,13}. Finally, SL *trans*-splicing, although not confined to parasitic organisms, is present in a wide range of medically and economically important parasites¹⁴. The presence of *trans*-splicing-specific factors indicates that this process is a possible target for therapeutic intervention. □

Methods

Extracts, splicing assays and RNA analysis

Ascaris lumbricoides whole-cell embryo extracts were prepared as described¹⁰. Mock depletion or immunodepletion was performed as described¹⁵ with antisera as indicated in figure legends. The *cis* splice substrate used was uniformly labelled with [α -³²P]GTP and corresponded to the wild-type *cis* substrate described in ref. 16. To assay bimolecular splicing, the same *cis* substrate was bisected within the intron; uniformly labelled 5' half-molecules were incubated with unlabelled 3' half-molecules. The 3' exon of the 3' half-molecule contains an exonic splicing enhancer, which promotes 3' splice site recognition in the absence of a 5' splice site, which is a prerequisite for bimolecular splicing¹⁶ (see also refs 7, 8).

To assay SL *trans*-splicing, 3' end-labelled (Fig. 2) or uniformly labelled synthetic SL RNA (Fig. 3) was incubated with the same 3' half-molecule used to assay bimolecular splicing. All splicing reactions were assembled, incubated and analysed as described¹⁶.

To identify RNAs associated with the M_r 30K and/or 175K proteins, immunoprecipitates of whole-cell extract obtained with either anti-175K or anti-30K polyclonal sera were deproteinized. Nucleic acids were recovered and end-labelled with [³²P]pCp as described¹⁶ before electrophoresis on denaturing gels. Digestion by RNase H with oligonucleotides complementary to the SL RNA was performed as described¹⁰.

Recombinant proteins

A full-length cDNA clone encoding the M_r 30K protein was fused in-frame with GST; the fusion protein was expressed in *Escherichia coli* and purified on glutathione–Sephadex. In some experiments (for example, that shown in Fig. 3c), the N-terminal GST tag was removed by cleavage of the fusion protein with thrombin. Similarly, fragments of SF1/BBP were expressed in *E. coli* as N-terminal GST fusion proteins and purified as above. For expression of the M_r 175K protein, full-length cDNA clones were cloned into a baculovirus vector such that the expressed protein contained a His₆ tag at its N terminus. The tagged protein was purified from whole-cell sonicates with Talon (Clontech) beads. To prepare 30K/175K heterodimers, the 175K protein was immobilized on Talon beads and exposed to a 10-fold molar excess of purified 30K protein. After being washed, the heterodimer was eluted with imidazole. Optimal amounts of recombinant proteins used for reconstitution were determined empirically by titration.

Received 22 January; accepted 26 March 2002; doi:10.1038/nature00783.

1. Blumenthal, T. & Steward, K. in *C. elegans II* (ed. Priess, J. R.) 117–146 (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1997).
2. Nilsen, T. W. in *Frontiers in Molecular Biology: Eukaryotic mRNA Processing* (ed. Krainer, A. R.) 310–334 (IRL, Oxford, 1997).
3. Bruzik, J. P., van Doren, K., Hirsh, D. & Steitz, J. A. *Trans*-splicing involves a novel form of small ribonucleoprotein particles. *Nature* **335**, 559–562 (1988).
4. Thomas, J. D., Conrad, R. C. & Blumenthal, T. The *C. elegans trans*-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. *Cell* **54**, 533–539 (1988).
5. Maroney, P. A., Hannon, G. J., Denker, J. A. & Nilsen, T. W. The nematode spliced leader RNA participates in *trans*-splicing as an Sm snRNP. *EMBO J.* **9**, 3667–3673 (1990).
6. Denker, J. A., Maroney, P. A., Yu, Y.-T., Kanost, R. A. & Nilsen, T. W. Multiple requirements for nematode spliced leader RNP function in *trans*-splicing. *RNA* **2**, 746–755 (1996).
7. Chiara, M. D. & Reed, R. A two-step mechanism for 5' and 3' splice-site pairing. *Nature* **375**, 510–513 (1995).
8. Bruzik, J. P. & Maniatis, T. Enhancer-dependent interaction between 5' and 3' splice sites in *trans*. *Proc. Natl Acad. Sci. USA* **92**, 7056–7059 (1995).
9. Abovich, N. & Rosbash, M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403–412 (1997).
10. Hannon, G. J., Maroney, P. A. & Nilsen, T. W. U small nuclear ribonucleoprotein requirements for nematode *cis*- and *trans*-splicing *in vitro*. *J. Biol. Chem.* **266**, 22792–22695 (1991).
11. Rain, J.-C., Rafi, Z., Rhani, Z., Legarin, P. & Krämer, A. Conservation of functional domains involved in RNA binding and protein–protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA* **4**, 551–565 (1998).
12. Nilsen, T. W. Evolutionary origin(s) of SL-addition *trans*-splicing: Still an enigma. *Trends Genet.* **17**, 678–680 (2001).
13. Stover, N. A. & Steele, R. E. *Trans*-spliced leader addition of mRNAs in a cnidarian. *Proc. Natl Acad. Sci. USA* **98**, 5693–5698 (2001).
14. Davis, R. E. Spliced leader RNA *trans*-splicing in metazoa. *Parasitol. Today* **12**, 33–40 (1996).
15. Makarova, O. V., Makarov, E. M. & Lüthmann, R. The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *EMBO J.* **20**, 2553–2563 (2001).
16. Maroney, P. A., Romfo, C. M. & Nilsen, T. W. Functional recognition of the 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol. Cell* **6**, 317–328 (2000).
17. Mazroui, R., Puoti, A. & Krämer, A. Splicing factor SF1 from *Drosophila* and *Caenorhabditis*: Presence of an N-terminal RS domain and requirement for viability. *RNA* **5**, 1615–1631 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank J. Bruzik, J. A. Wise and M. Caprara for comments on the manuscript, and A. M. Micenmacher for help with figures and manuscript preparation. The work was supported by grants from the N.I.H.

Competing interests statement

The authors declare that they have no competing financial interests

Correspondence and requests for materials should be addressed to T.W.N. (e-mail: tw@po.cwru.edu).

DNA and protein microarrays

Microarrays are everywhere, doing everything...

Kv subunit protein array

Sense Proteomic www.senseproteomic.com

Cell-free screening with potential

Sense's Kv subunit protein array is a collection of functional potassium channel regulatory proteins immobilized in microtitre plate format for use with conventional laboratory equipment. It contains key human and rat potassium channel β -subunit proteins. The channel subunits can be interrogated to identify interacting protein pathways and provide a cell-free method of screening new chemical entities for effects on channel-protein interactions and their potential to modulate potassium channel function.

FluoroScript

Invitrogen www.invitrogen.com

Microarray screening

Invitrogen's FluoroScript system provides fluorescent cDNA targets using a streamlined protocol. FluoroScript RT, an avian RNase H⁻ reverse transcriptase, is used to achieve incorporation of labelled nucleotides (10–20%) with high cDNA yield and length.

HyPro₂₀/HyPro₁₀₀

Thermo Life Sciences www.thermols.com

Molecules in motion

HyPro₂₀ and HyPro₁₀₀, based on the OmniSlide *in situ* PCR system, provide batch microarray hybridization of up to 20 or 100 slides. An active vibration system encourages increased molecular movement of the solution in contact with the immobilized probe, eliminating local depletion of hybridizing target at the microarray platform surface. A flat heating block minimizes inter- and intra-slide variability. HyPro can be used with cover slips or adhesive hybridization chambers and, at approximately 25 × 75 mm, is suitable for use in most slide hybridization assays.

EURAY Immobilizer

Flowgen www.flowgen.co.uk

Rooted to the spot

This polymer microarray slide has a transparent spotting area coated with probes that enable researchers to perform 'on-the-spot' amino-linked DNA/RNA, -LNA or -PCR spotting. Between the spots there is low, non-specific binding of DNA and fluorochrome labelled proteins. The slides require no activation steps, blocking, baking or UV-cross-linking and are compatible with most microarray spotters and scanners. Spotting and hybridization solutions are provided with the slides.



Spot on: Flowgen's polymer microarray slide.

Applications include clone mapping, linkage analysis, highly multiplexed mutation detection, gene expression monitoring and high-throughput screening.

RNA 6000

Agilent Technologies www.agilent.com

A chip off the old block

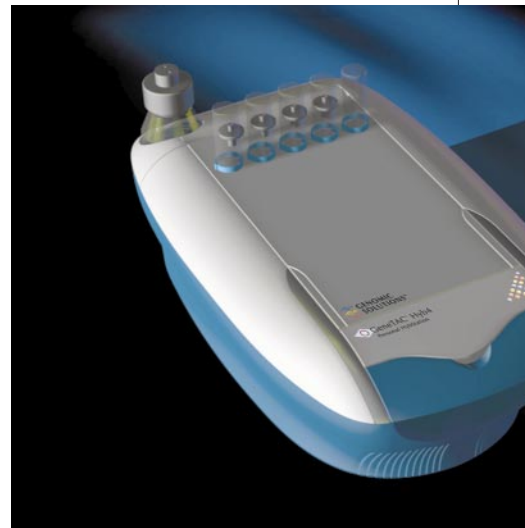
Designed for use with Agilent's 2100 bioanalyzer, the RNA 6000 Nano LabChip kit is an updated version of the previous RNA 6000 kit. The new kit contains a lower marker that enables researchers to distinguish different types of mRNA based on electrophoretic traces. RNA quantification is also improved, and can be carried out simultaneously with assessment of RNA quality. Ultraviolet or ribogreen measurements are not necessary in many cases, with quality control and concentration measurement performed in a single analysis. Relative standard deviations of 10% can be achieved. The kit allows detection of total RNA down to 5 ng.

Absolutely

Stratagene www.stratagene.com

RNA microprep, pure and simple

The Absolutely kit from Stratagene is designed for isolating total RNA from small starting samples and laser microdissection samples containing 1 cell to 5×10^5 cells. Isolation of total RNA takes place in a single spin cup, with DNA removal performed directly in the cup to maximize RNA recovery. The resulting RNA can be used in applications such as purification of laser microdissection samples for quantitative RT-PCR and microarray target generation.



Change TAC: the compact GeneTAC Hyb4.

GeneTAC Hyb4

Genomic Solutions www.genomicsolutions.com

For the researcher working on all fours

Genomic Solutions' four-position GeneTAC Hyb4 is a compact unit for individual researchers who need to run fewer than 12 microarrays at a time. It automates all hybridization temperature cycles and post-hybridization washes of microarrays spotted on standard glass slides. The unit comes with a laptop computer and multi-protocol software to enable the user to run multiple units from the same computer for independent sets of hybridizations, or to run all arrays under the same conditions. A universal slide carrier that accommodates four slides is also included.

DNA Microarray Expression Analysis

Integrated Genomics

www.integratedgenomics.com

Pathfinding service

Integrated Genomics offers a choice of customized microarray design, full-line microarray manufacturing or off-the-shelf microarrays. Experimental support includes normalization and statistical analysis of microarray data and expression data analysis in the context of networked pathways to help identify functions for uncharacterized open reading frames (ORFs). A sequenced and annotated genome provides the sequence, possible ORFs and their potential function. Full-genome arrays for *E. coli*, *S. aureus*, *B. subtilis* and *F. nucleatum* are available. Custom-ordered microarrays are designed to have known sequences of similar length representing each gene in the genome. Each spot included on the microarray represents one ORF.

DNAnchor

Synthetic Genetics www.syntheticgenetics.com

Form a lasting attachment

DNAnchor slides and oligonucleotides use terminal attachment chemistry designed to produce higher signal intensities for constructed DNA microarrays. The three-step process is performed in about three hours. Each kit contains five coated slides, a hybridization container, three bottles of wash buffer and one microtube of 20X attachment buffer. After printing, the slides are incubated for one hour. Covalent coupling of the modified probe to the slide is then complete, without the need for a reduction step. Applications include cDNA expression microarrays, oligonucleotide microarrays and SNP typing.

Expression Profiling Service

GeneScan www.genescan.com

Biochip-based gene activity profiles

GeneScan Europe AG is offering a comprehensive, modular service for investigating gene activity profiles on the GeneScan biochip platform. This Expression Profiling Service includes all the experimental stages involved in a biochip-based investigation of gene activity profiles. The microarrays are either specially designed for the customer as ScienceChips or are already available as finished application chips. All investigative stages, from sample preparation, two-colour analysis on the chip, readout of the chips to data evaluation via an interactive 'data bridge', are available in the form of flexible service modules.

Micromax

PerkinElmer Life Sciences

www.perkinelmer.com/lifesciences

Microarrays for apoptosis and neurobiology

These new glass-based microarrays — one for apoptosis and one for neurobiology applications — have smaller gene sets immobilized on them, enabling researchers to conduct more focused, cost-effective procedures. Human apoptosis (MPS606) and human neurobiology (MPS607) microarrays are additions to the Micromax microarray product family. The apoptosis microarray contains 333 genes plus 27 control genes, and the neurobiology microarray contains 402 neurobiology genes plus 32 control genes.

Template preparation kits

Roche www.roche-applied-science.com

Three new ways to hit the target

Three template preparation kits from Roche have been optimized specifically for microarray applications. The microarray target amplification kit requires only 50 µg tRNA and performs cDNA synthesis in a one-tube reaction. The target synthesis kit (T7) uses only a small amount of cDNA (from 1 µg

RNA) and is compatible with a broad spectrum of hapten- and fluorophore-labelled nucleotides. A yield of 100 µg is produced in a reaction time of only two to three hours. Finally, the purification kit has been designed to purify a variety of synthesized nucleic acids. It removes substances that would interfere with highly sensitive hybridization. A low-maintenance protocol minimizes handling.

MessageAmp

Ambion www.ambion.com

Is this the end of home-made Eberwine?

The MessageAmp aRNA kit selects and amplifies mRNA from total RNA or poly(A) samples, resulting in a 1,000-fold amplification of the RNA target for array analysis. The kit is based on Eberwine's procedure and is compatible with both biotin and Cy dye labelling. Reagents optimized for each step of the reaction are included.

GenoSensor Array 300

Vysis www.vysis.com

Screen multiple gene targets

The Vysis GenoSensor system allows assessment of multiple gene targets for identification of gene amplifications or deletions in a variety of different sample types. GenoSensor offers a complete approach to genomic research, and includes genomic arrays, instrumentation and reagents using Vysis' patented comparative genomic hybridization (CGH) technology. It is able to screen multiple gene targets in a single test.

ExpressArray

Digene www.digene.com

Hybrid capture array kit

Digene's new hybrid capture assay allows researchers to run highly complex tests involving the study of thousands of genes in a single experiment. Quantitative results are generated in hours rather than days. The kit requires no target labelling or amplification. Improved sensitivity allows users to start with less than one microgram of total RNA.

Transignal

Panomics www.panomics.com

Multiple TF characterization

Panomics has introduced the TranSignal protein DNA array for high-throughput functional analysis of eukaryotic transcription factors (TFs). The company describes this new technology as a significant improvement over gel mobility-shift assays, which permit characterization of only one TF at a time. The TranSignal array characterizes the activities of 54 TFs simultaneously. When the TranSignal probe mix is added to nuclear extracts from the two samples being compared, DNA/protein complexes form, representing the TFs of interest. Bound probes are separated from unbound probes, and the formerly TF-bound DNA sequences are hybridized to array membranes. The signal is detected by chemiluminescence. No radioactivity is necessary.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Fabien Savenay

Editor: Paul Smaglik

Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Senior European Sales Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Director: Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.johnson@naturejp.com

Asia-Pacific Advertising Manager:

Kevyn Johnson

naturejobs

Landing the egos

In helping to start three biotech companies in the Seattle area (see pages 4–5), Chris Henney has had to be a good judge of scientific talent. But he calls himself “a pretty bad interviewer”. So how has he managed to recruit scientists for Immunex — which is in the process of being acquired by Amgen of Thousand Oaks, California — Icos, which hopes to market its first drug in a year; and Dendrion, which is still developing? “I usually just say: ‘Tell me what you’ve been doing,’” he says.

This open-ended question usually reveals more than people intend. “I look for high-energy, open-minded people,” Henney says. What does he want to avoid? “Guys who want to spend their life repeating their PhD thesis,” he says. Shaking his head, he adds in his staccato mantra: “It’s probably not going to work.”

The other thing he watches for — both in interviews and once a new recruit has started — is ego. “We all have well-developed egos in the biotech world,” Henney says. But he emphasizes that a little goes a long way, and that the line between confidence and arrogance can be relatively blurred.

So what are some ego-related disqualifiers? People who worry about other people going into their office. “It’s probably not going to work,” says Henney. People who want their own private parking space. “It’s not gonna work,” says Henney. A desk with a name embossed in metal or some exotic polymer. “Not gonna work,” he says.

The best kind of confidence, according to Henney, is the kind that has to do with honest intellectual pursuit, rather than the trappings of success or status. “You’re looking for people who don’t have to win every argument,” Henney says. That qualification is desirable in any scientific setting, academic or industrial.

Paul Smaglik
Naturejobs editor



Contents

REGIONS

Seattle and Vancouver make
their mark in biotech

p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

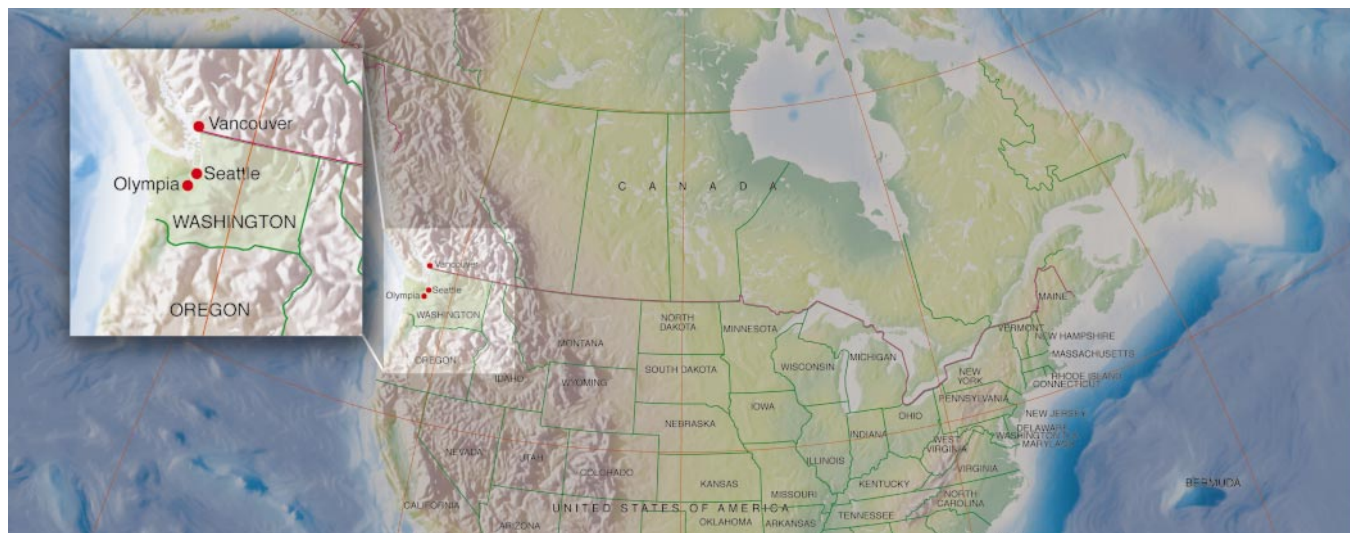
SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS





The best of both worlds Seattle

Seattle has always been a boom-and-bust kind of town. Boeing's success made it an aerospace capital — until the company decided to move its corporate headquarters to Chicago. The software dominance of Microsoft and the rise of the Internet spurred hundreds of dotcom companies — until the bubble burst last year. Now, with local unemployment higher than the national average, the city is turning tentatively to what it hopes is the next big thing.

"Everyone is looking at biotech to be the next saviour," says Ruth Scott, president of the Washington Biotechnology and Biomedical Association. Actually, she acknowledges that biotech has had a long-standing presence in the Pacific Northwest, but it has been relatively quiet; the area lacks 'big pharmaceuticals'. Now the Seattle biotech scene seems to have reached a turning point.

One mark of this is the change in ownership of two of the region's most successful companies. Rosetta Inpharmatics was acquired by Merck last July, and Immunex, the city's biggest biotech company, is being bought by Amgen.

Seattle's biotech hub: the University of Washington (below) and the Hutch (right).



Rosetta's experience may offer some insight into what the future holds for Immunex as Amgen moves in. In some ways, Rosetta is the archetype of a Seattle-based biotech company. It was built on technologies developed at the two main local research institutions: the University of Washington and the Fred Hutchinson Cancer Research Center (known as 'the Hutch').

SILICON ROOTS

Both technologies had some computational genesis: a technique for printing gene-expression arrays from Leroy Hood's University of Washington lab and a bioinformatics approach to analysing expression data from Lee Hartwell's lab at the Hutch. Rosetta was also partly funded by Vulcan Ventures, a venture-capital fund operated by Paul Allen, one of Microsoft's founders.

Lisa Junglov, Rosetta's recruiting manager, says that since Merck acquired the company, it has grown from 180 employees to about 230. Now Rosetta gets lots of unsolicited applications from unemployed IT people, when what it really wants is molecular biologists with some computational skills.

But for younger companies such as Illumigen Biosciences, this climate might be just right. It hired four IT people in its early days — three of whom had been laid off from Internet start-ups. "To a certain degree, the collapse of the dotcom bubble has helped people like us," says Shawn Iadonato, the firm's chief scientific officer.

The company has also been aided by input from two people who played a huge role in sequencing the human genome. Illumigen emerged from

Vancouver's widening horizons

Three graduates of Simon Fraser University (SFU) who left Canada for world-renowned genomics centres are now in leadership positions in Vancouver. They represent early success in reversing Canada's brain drain (see *Naturejobs* 4–5; 14 February 2002).

Francis Ouellette, Steve Jones and Marco Marra are helping the city position itself as a bioinformatics hub in the Can\$300-million (US\$195-million), three-year Genome Canada programme. Genome British Columbia, one of five regional divisions, will get roughly a fifth of that money, with much of it concentrated near Vancouver.

Ouellette spent time at the US National Center for Biotechnology Information before becoming bioinformatics

director at the University of British Columbia's Center for Molecular Medicine and Therapeutics. He has accepted the directorship of a new bioinformatics centre at the university.

Jones left SFU for the Sanger Center in Britain, and is now running Canada's newest graduate-level bioinformatics training programme, at the British Columbia Cancer Agency's Genome Sciences Centre (GSC).

And Marra, director of the GSC, sees even more opportunities appearing, because funding from the Cancer Agency and support from the US National Institutes of Health are abundant. Recruitment emphasizes bioinformatics, gene expression, genotyping and SNP discovery.

P.S.

Maynard Olson's University of Washington lab, which led the university's part in sequencing the human genome. And Philip Green, who provided the Human Genome Project with bioinformatics expertise, also sits on the company's scientific advisory board.

A more visual example of the dynamics between IT and bioinformatics can be found at the Institute for Systems Biology (ISB), which Hood left the University of Washington to establish two years ago. The institute is in a waterfront building that once housed eight dotcom companies — none of which are still in business.

Hood, like Iadonato, has been able to find IT people in the wake of the collapse, although “probably not enough”, he says. The institute has grown from 2 to 170 people in 18 months, and has nine faculty members including chemists, engineers, mathematicians and even an astrophysicist. Hood anticipates that, in addition to more computational biologists, his institute will need more mathematicians in order to flourish.

There are also signs that Immunex's merger may help the whole Seattle biotech area to flourish in the long term, although the short-term indications are less positive. On the downside, the company has drastically scaled back plans for a new research facility and there are rumours that some employees are nervous they may be relocated to Amgen's site in Thousand Oaks, California.

But Immunex has history on its side. It originally emerged from the labs of scientists at the Hutch. And since its inception, it has sprouted several other companies. For example, Immunex co-founder Chris

Henney has since created Icos, which expects to market a Viagra-like drug within a year, and Dendreon, where he currently acts as chairman and chief executive. Both companies are hiring scientists now.

He is in the camp — along with Ruth Scott and others — that believes more companies will emerge from Immunex after the merger. Although the area does not have the venture-capital market of Boston or the Bay, “there's a lot of money in this area which is very Northwest-oriented”, he says.

SPIN CITY

Targeted Genetics, yet another Immunex spin-off, is demonstrating how its parent company has been an engine to produce yet more — it is planning to spin out its cell-therapy unit into a company called CellExSys. Perhaps one of the reasons for so many local spin-offs is the reluctance of their employees to go elsewhere. Geoff Roach, senior director of human resources at Targeted Genetics, calls that phenomenon the “emerald handcuffs”. People who like the outdoors but still want to live in or near a city are attracted to Seattle, he says.

Newer institutions are also taking advantage of the area's amenities. The ISB has a deck equipped with barbecue grills overlooking Elliot Bay, and ZymoGenetics, a biotech firm in a renovated power plant, beats that by offering its own dock — and employee access to rowing shells.

Don Elmer, managing general partner at Pacific Horizons Ventures, has his own perspective of the scene. Unlike most venture-capital firms, which tend to be located close to areas of dense activity, Elmer's firm follows companies that are quite widely spread out. But this is not a problem, Elmer says, and the reason is environmental. One partner at the firm delights in the scenery he is presented with as he travels to the more far-flung institutes in Montana or British Columbia. Elmer jokes that the approach is akin to the Hudson Bay fur traders making the rounds for beaver pelts.

What he and his partners have found during those rounds is cause for optimism. They are seeing signs that the area's convergence of IT and biotech is creating an improved ability to decide whether to pursue a particular drug candidate or target, thus avoiding “late-stage clinical-trial fiascos”. The ability to save money this way is “the promise of computational biology”, he says. And the hope of the Seattle biotech scene. ■

Paul Smaglik is editor of *Naturejobs*.

Washington Biotechnology & Biomedical Association

♦ www.wabio.org

Genome British Columbia

♦ www.genomebc.ca



Leroy Hood: seeking mathematicians.



Biotech pioneers: Chris Henney (right) with Martin Simoneti (centre) and David Urdal.